

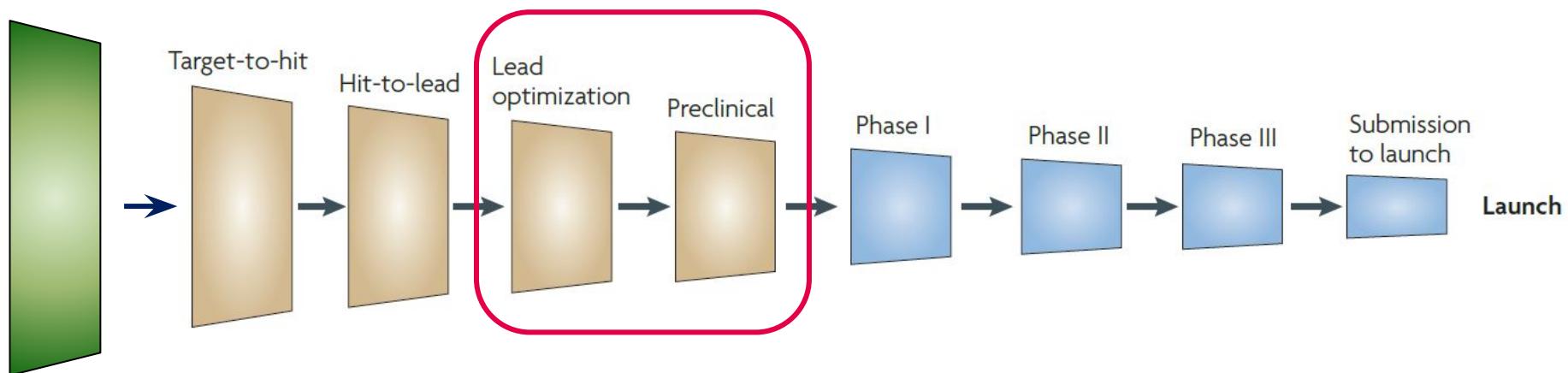
# What efficacy and safety profiles can we expect

*Mathematical and Computational Biology in Drug Discovery  
(MCBDD) Module IV*

*Dr. Jitao David Zhang  
May 2024*

# Where are we now

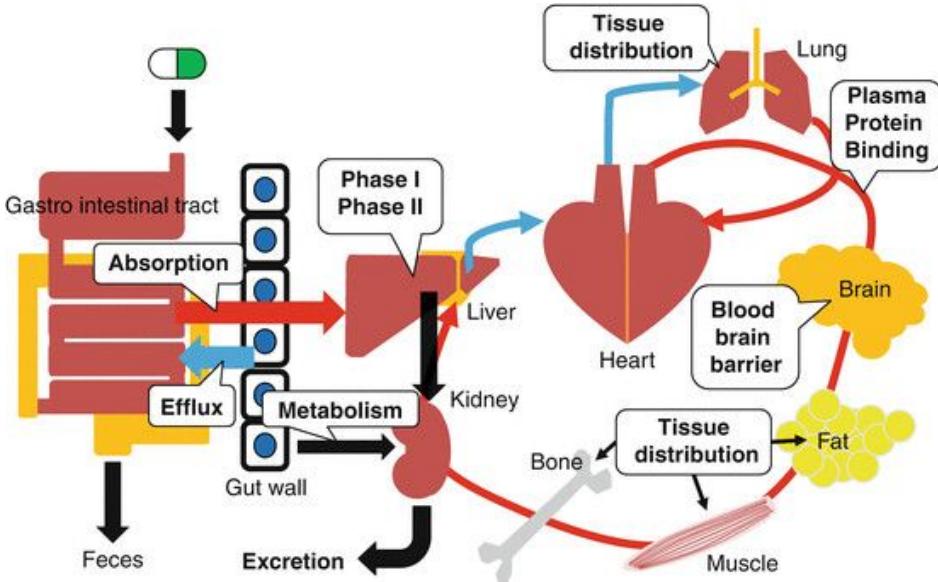
Target identification & assessment



**Goal:** we want to select **one compound** from a few (~ $10^2$ - $10^0$ ) for entry in human.

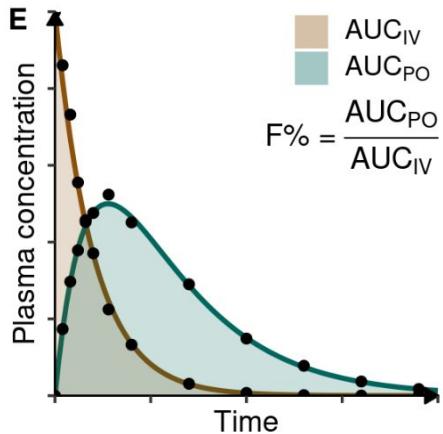
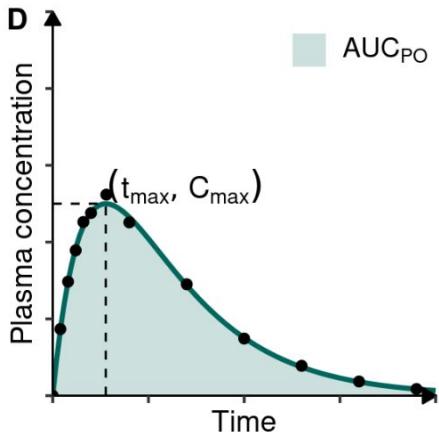
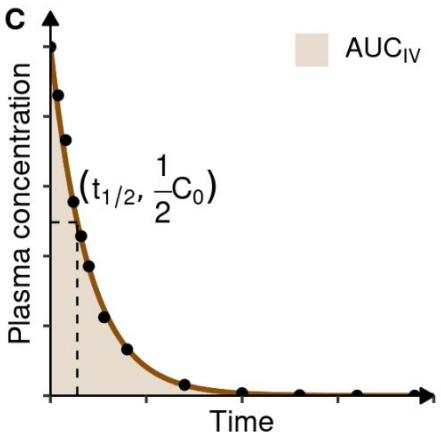
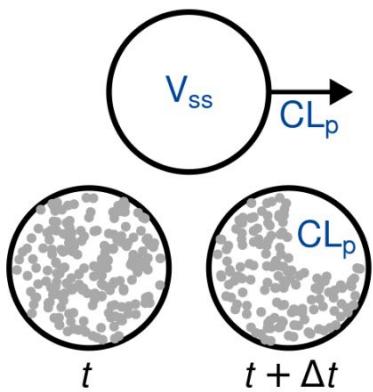
# Key factors to consider in selecting compounds

- Efficacy and pharmacodynamics(PD)
- Pharmacokinetics (PK)
  - Absorption
  - Distribution
  - Metabolism
  - Excretion
- Toxicology



# Key PK parameters: V<sub>ss</sub>, CL<sub>p</sub>, t<sub>1/2</sub>, t<sub>max</sub>, C<sub>max</sub>, and F

B



|                        |   |                        |   |
|------------------------|---|------------------------|---|
| <b>V<sub>ss</sub></b>  | Volume of distribution at steady state.   | <b>C<sub>max</sub></b> | Maximum plasma concentration.   |
| <b>CL<sub>p</sub></b>  | Plasma clearance.   | <b>t<sub>max</sub></b> | Time point in which the Cmax is measured.   |
| <b>t<sub>1/2</sub></b> | Half-life, time for a substance to reach the half concentration of the initial value (C <sub>0</sub> ). | <b>F% (or F)</b>       | Bioavailability, the percentage of the administered compound reaching systemic circulation. |

# The Hill function as a typical PD model

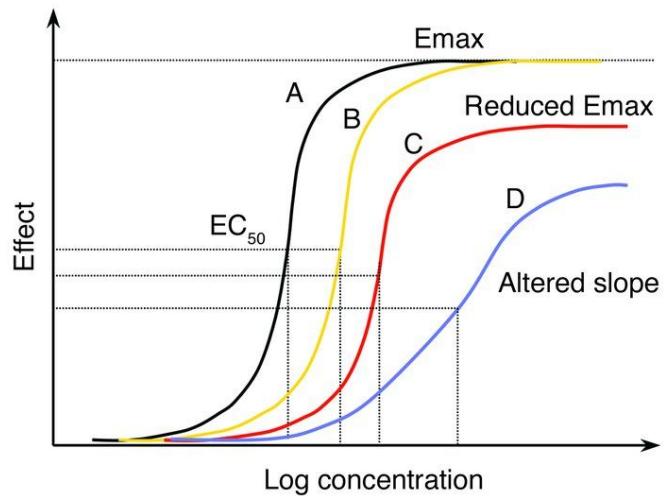
- The Hill function is one of the mostly useful non-linear functions to model biological systems.
- In its general form,  $H_{max}$  indicates the maximal value to which the function is asymptotic,  $n$  is the shape parameter (known as the Hill's coefficient), and  $k$  is the reflection point, often abbreviated as  $XC_{50}$  ( $X=I, E, C, \dots$ ), the half-saturation constant.
- The Michaelis-Menten model is a special case of the Hill function with  $n=1$ .

$$H = H_{max} \frac{x^n}{k^n + x^n}$$

**General form of the Hill function**

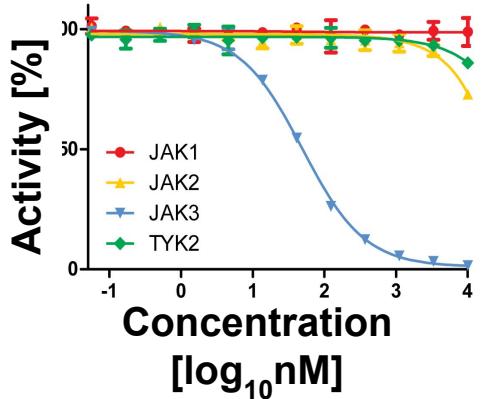
$$\begin{aligned} E &= E_{max} \frac{[L]^n}{EC_{50}^n + [L]^n} \\ &= E_{max} \frac{1}{1 + \left(\frac{EC_{50}}{[L]}\right)^n} \end{aligned}$$

**Modelling dose-dependent effect**

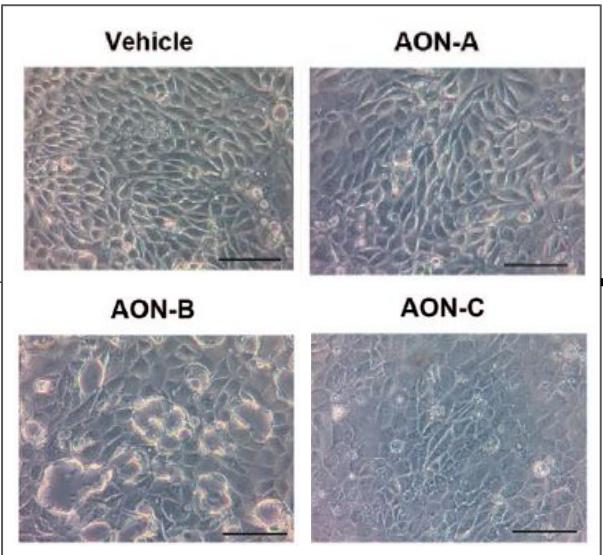


White. J Clin Invest. 2004;113(8):1084-1092.  
<https://doi.org/10.1172/JCI21682>.

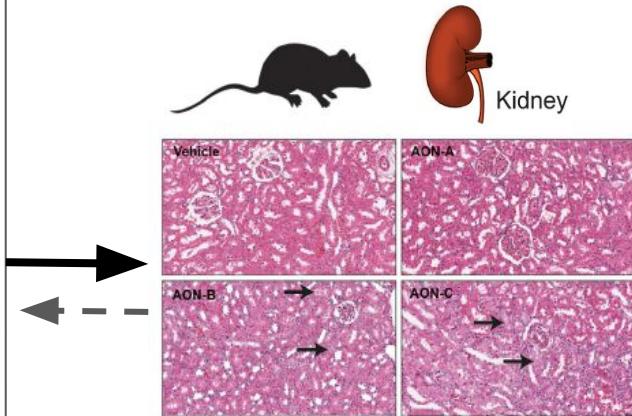
# Classical workflow of efficacy and toxicity assessment



Biochemical &  
biophysical assays



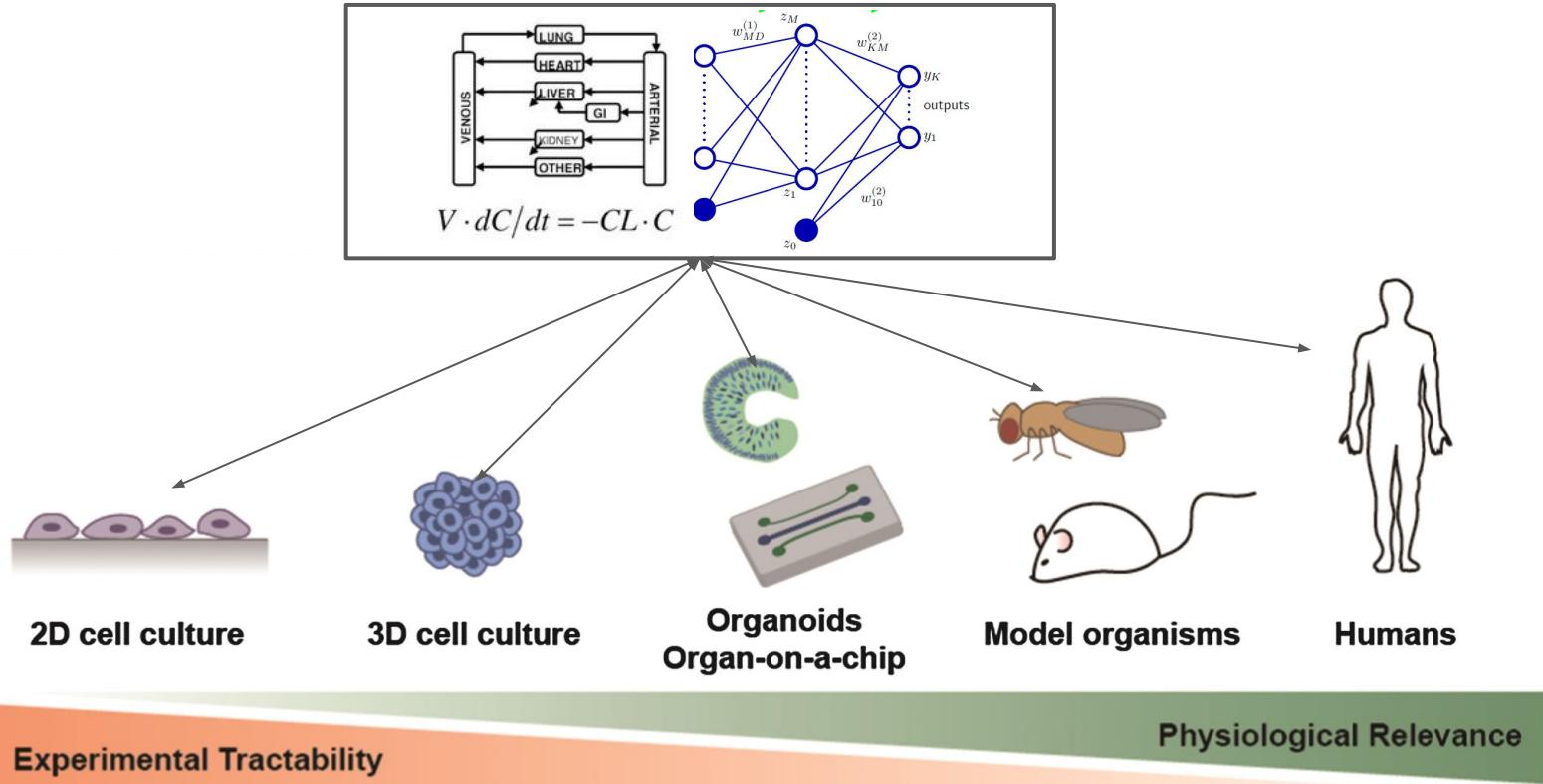
Cellular assays  
(*in vitro*)



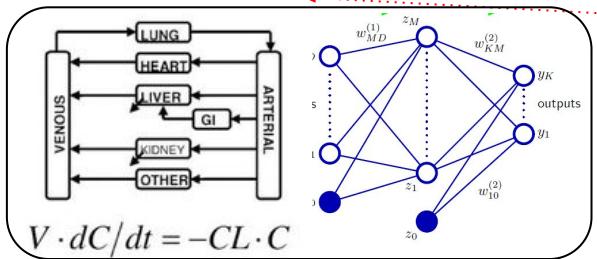
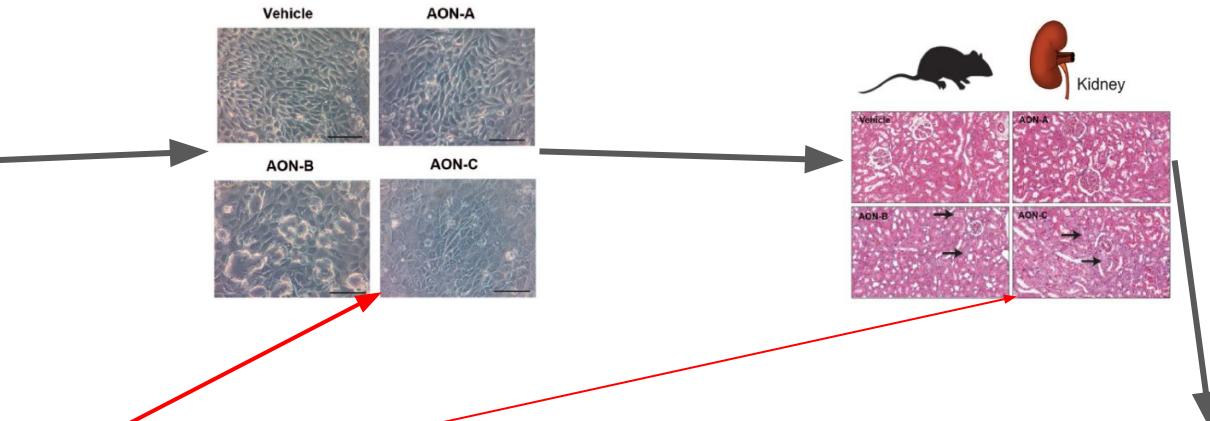
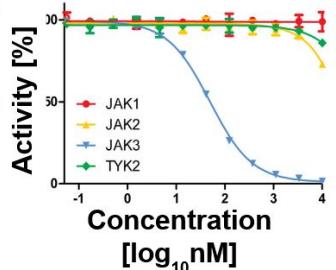
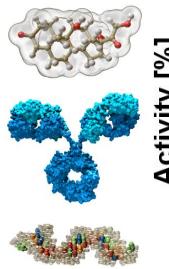
Animal  
experiments  
(*in vivo*)

→ Usual workflow  
 ← - - - Assay development

# Biological and computational models of human diseases



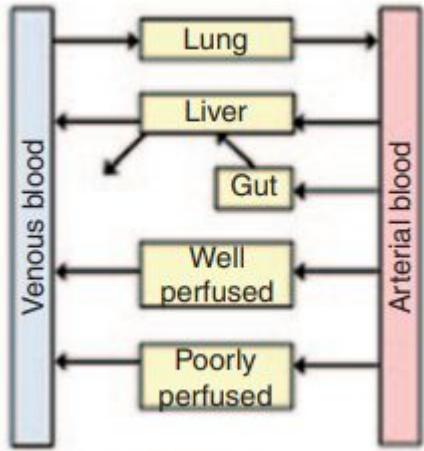
# Computational methods empower efficacy and toxicity assessment



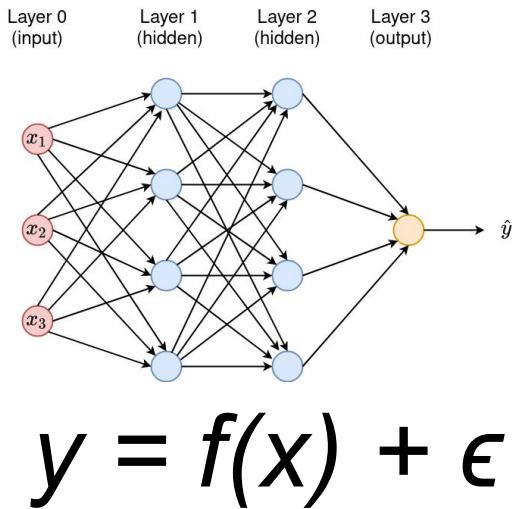
Mechanistic, causal,  
and statistical models



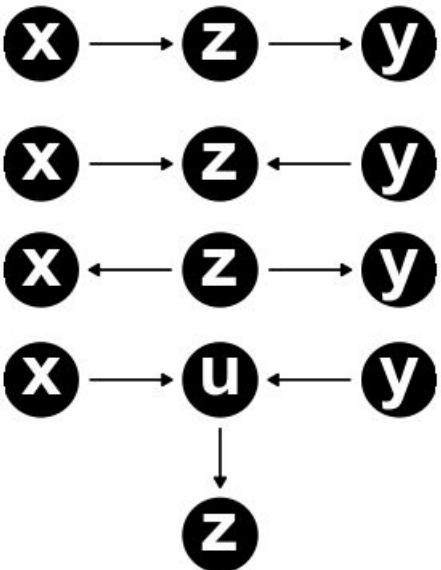
# Three types of computational models



Mechanistic models

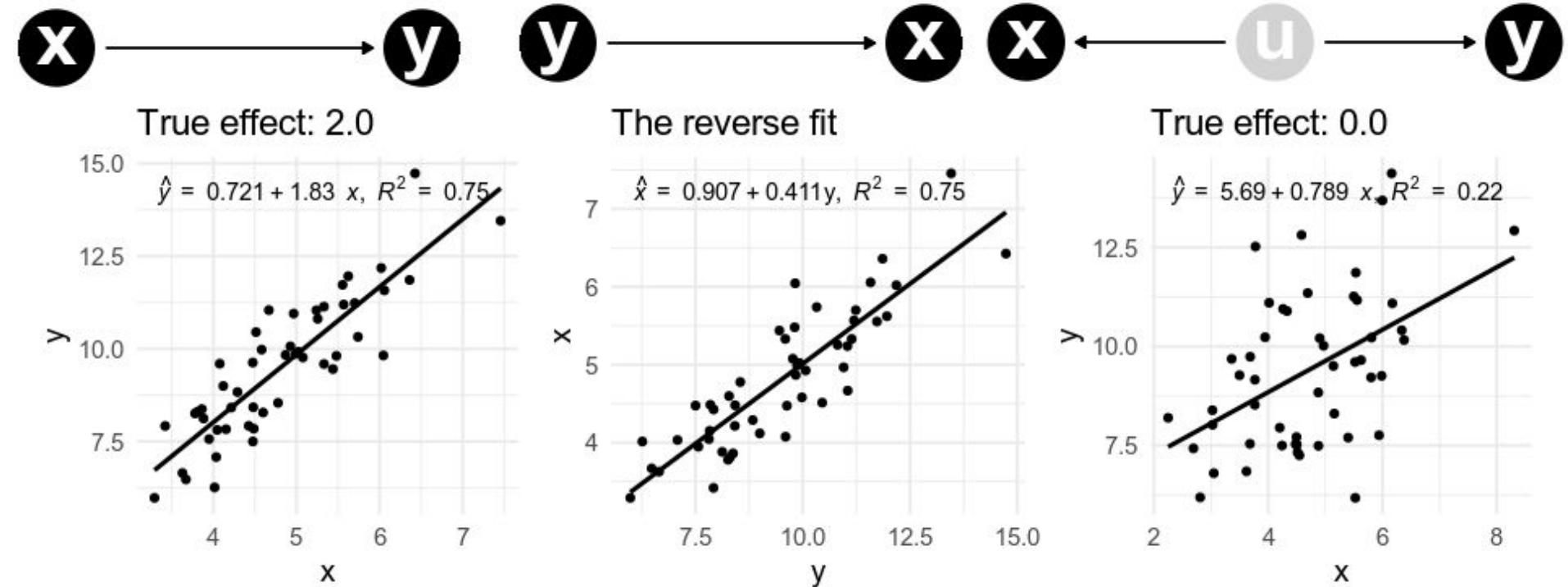


Statistical and  
machine-learning models



Causal models

# Correlation is caused by causation, confounding, coincidence, or conspiracy



Statistical models alone cannot derive causality from correlation

# We learn causality by (1) listing models explicitly and (2) manipulating a variable and observe the outcomes

**Model 1**



**Model 2**

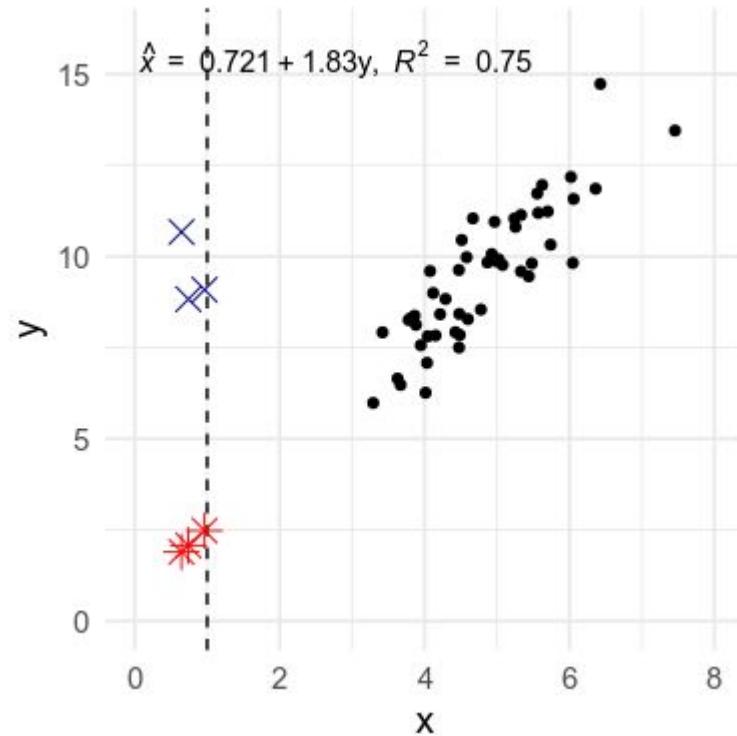


**Model 3**

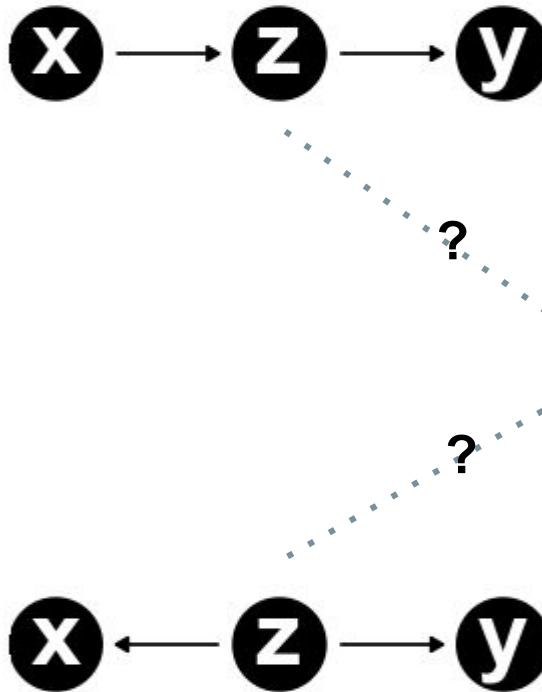


Assume that the data is generated by either Model 1, or Model 2, or Model 3. And assume that we can manipulate the value of X by setting it to 1.0 (the dash line).

Question: which outcomes (red stars or blue crosses) would support which models? Why?

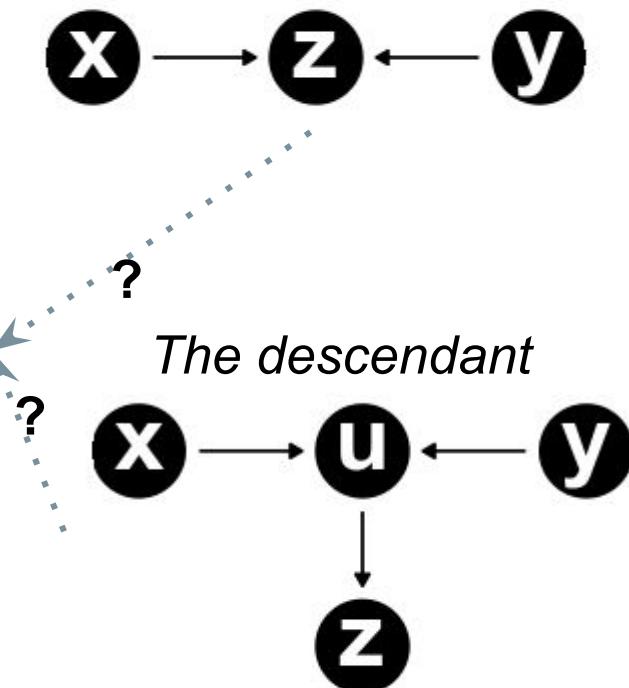


# Causality is crucial for drug discovery



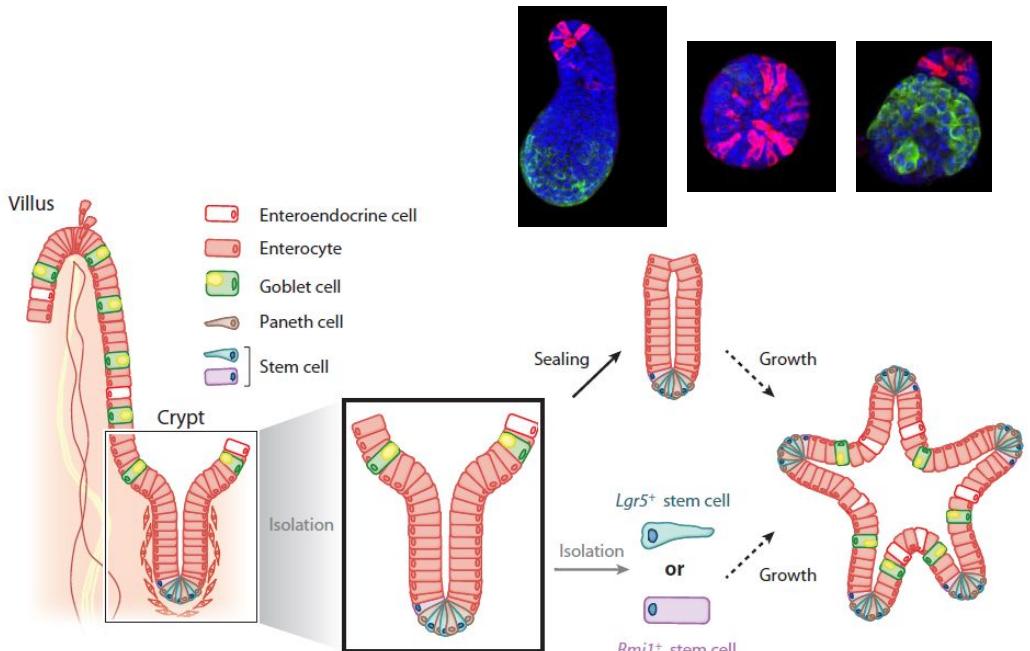
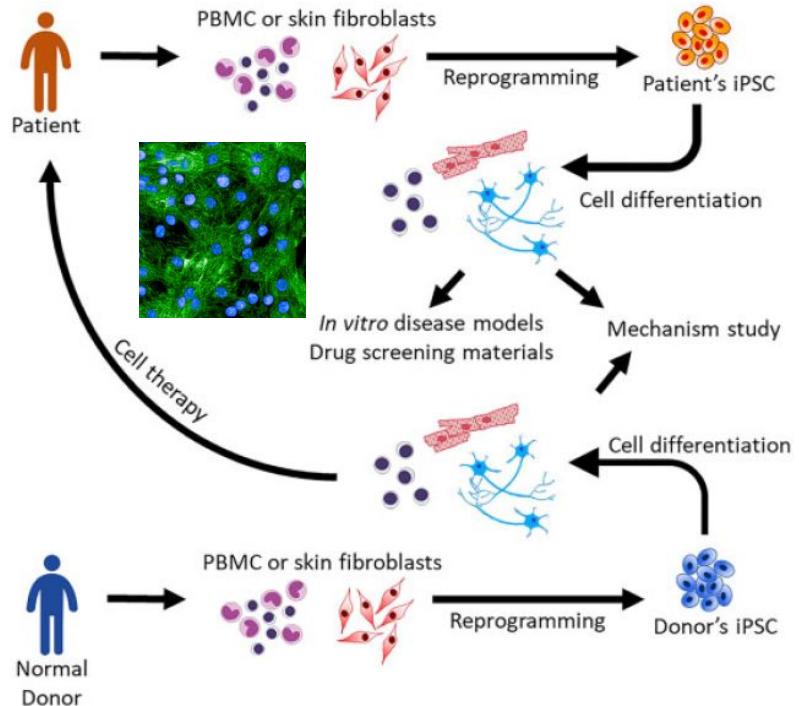
Biomarker, tox study, pathology,  
omics data, real-world data, ...

|    | x            | z  | y           |
|----|--------------|----|-------------|
| 1  | 0.835386320  | 1  | -0.73897252 |
| 2  | -0.005354014 | -1 | -0.82972315 |
| 3  | 0.058788286  | 1  | 0.76213369  |
| 4  | -1.015602246 | -1 | -0.05951719 |
| 5  | -0.339569780 | -1 | -0.11745910 |
| 6  | -0.041077979 | -1 | -1.28243716 |
| 7  | 0.363740407  | 1  | -0.30570762 |
| 8  | 0.119496314  | -1 | -1.19932461 |
| 9  | 0.257108454  | -1 | -1.06044066 |
| 10 | 0.304537158  | -1 | -0.43396492 |



We need both models (knowledge + assumptions) and data to infer causality.

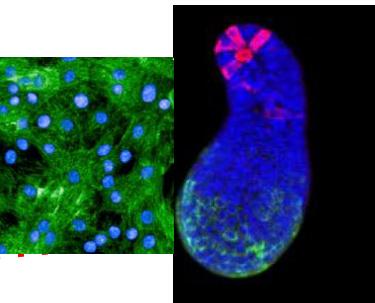
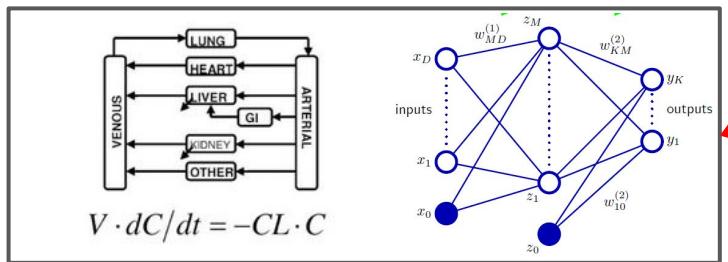
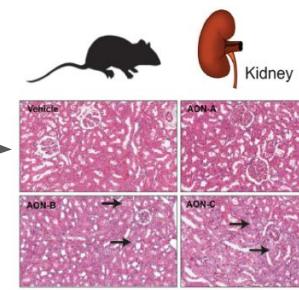
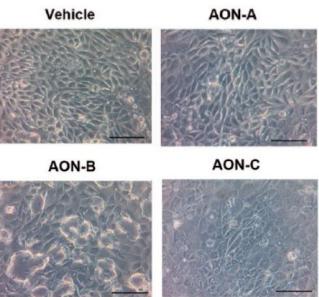
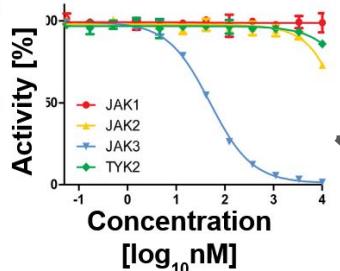
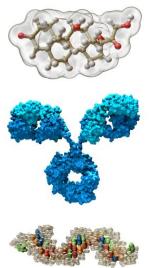
# Stem cells and organoids empower efficacy and toxicity assessment



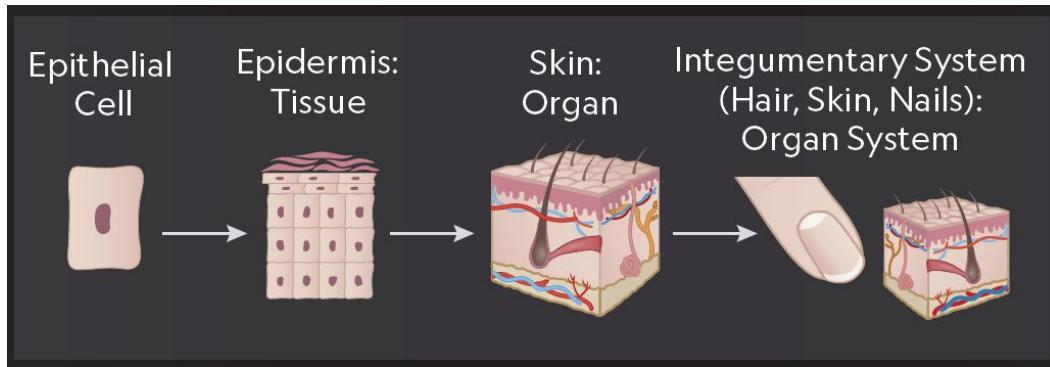
## Small-intestinal organoids

### Induced pluripotent stem-cells

# Computational methods and novel biological models empower efficacy and toxicity assessment



# Complexity Increases Through a System

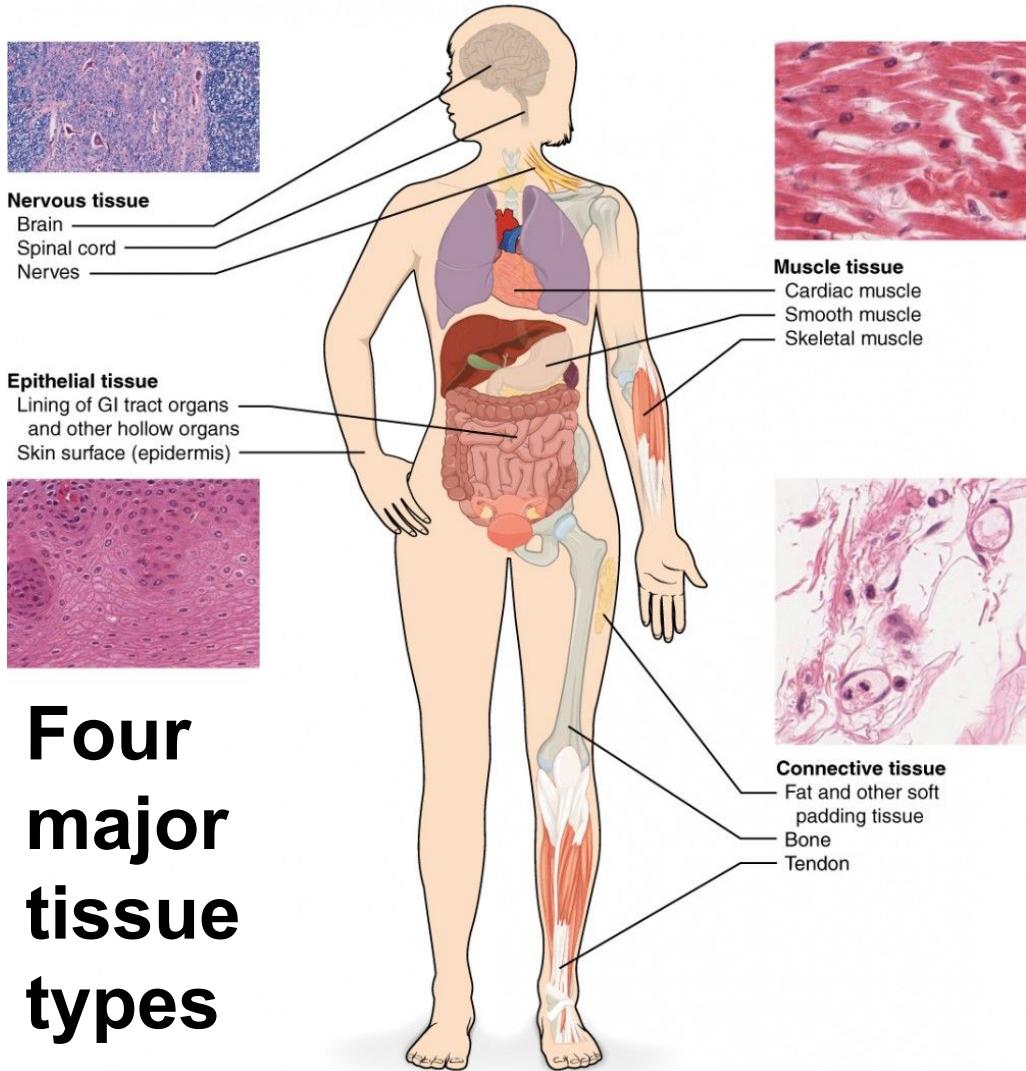
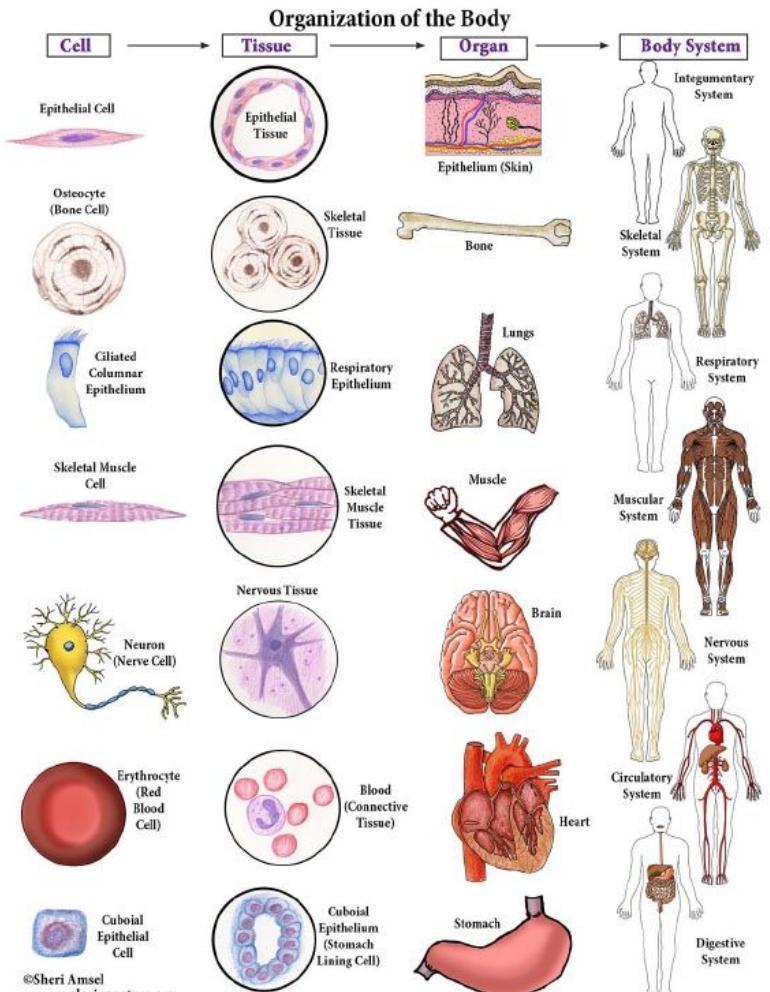


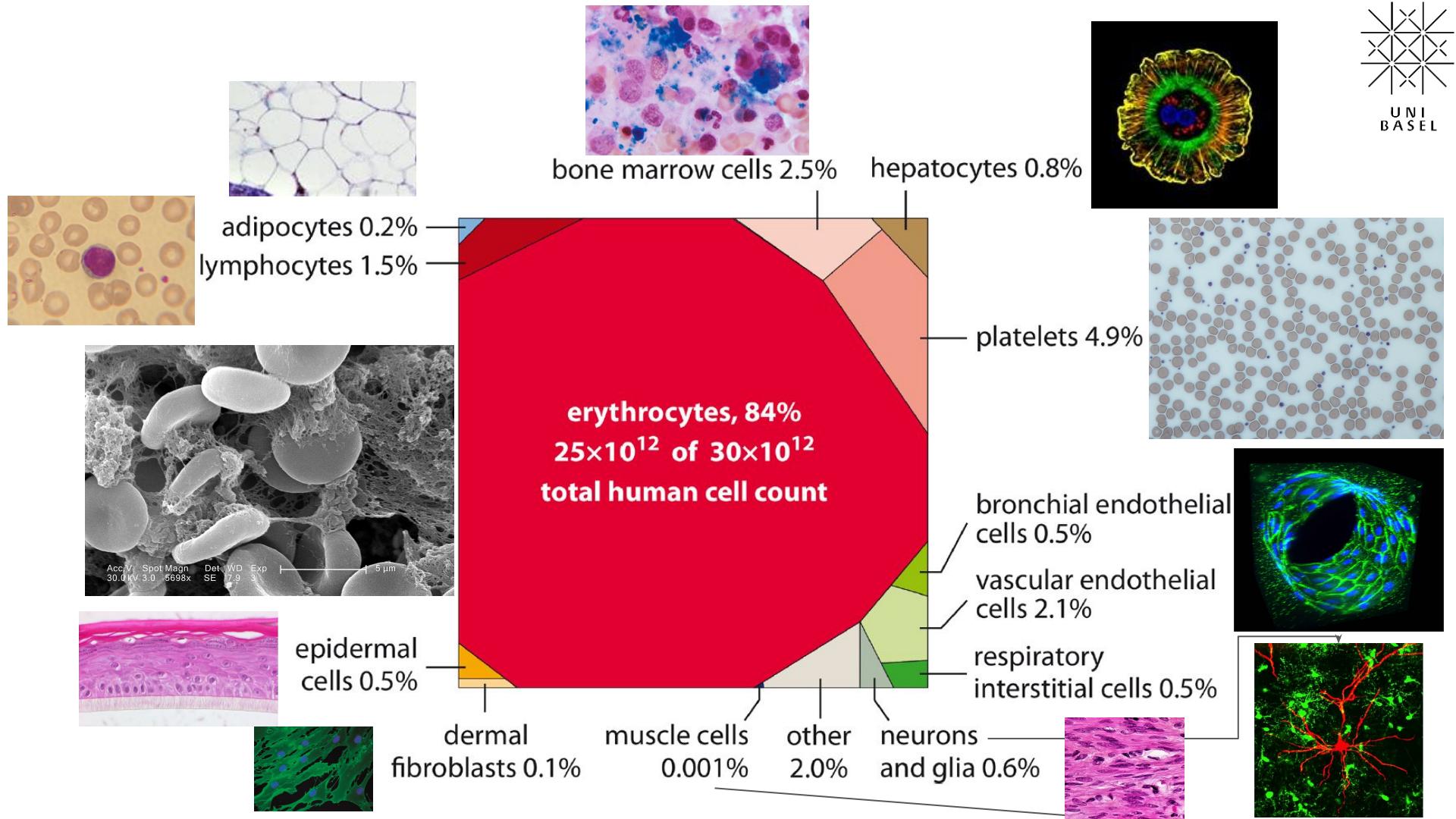
**Cells:** basic building blocks, variable morphologies and functions

**Tissues:** groups of specialized cells that communicate and collaborate

**Organ:** group of tissues to perform specific functions

**Organ systems:** group of organs and tissues





# What's in a drop of blood? Ask a doctor or a biologist!

**Plasma:**

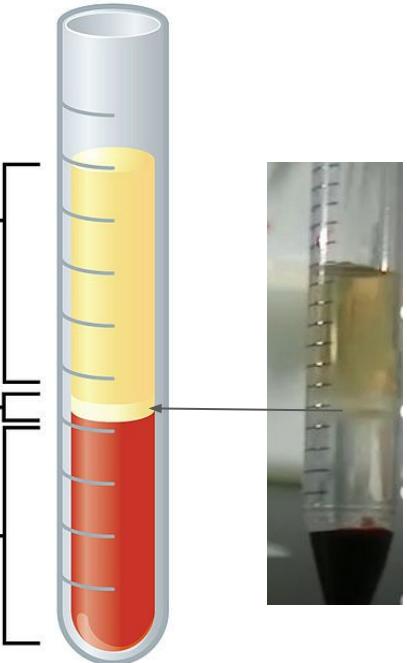
- Water, proteins, nutrients, hormones, etc.
- ~55%

**Buffy coat:**

- White blood cells, platelets
- <1%

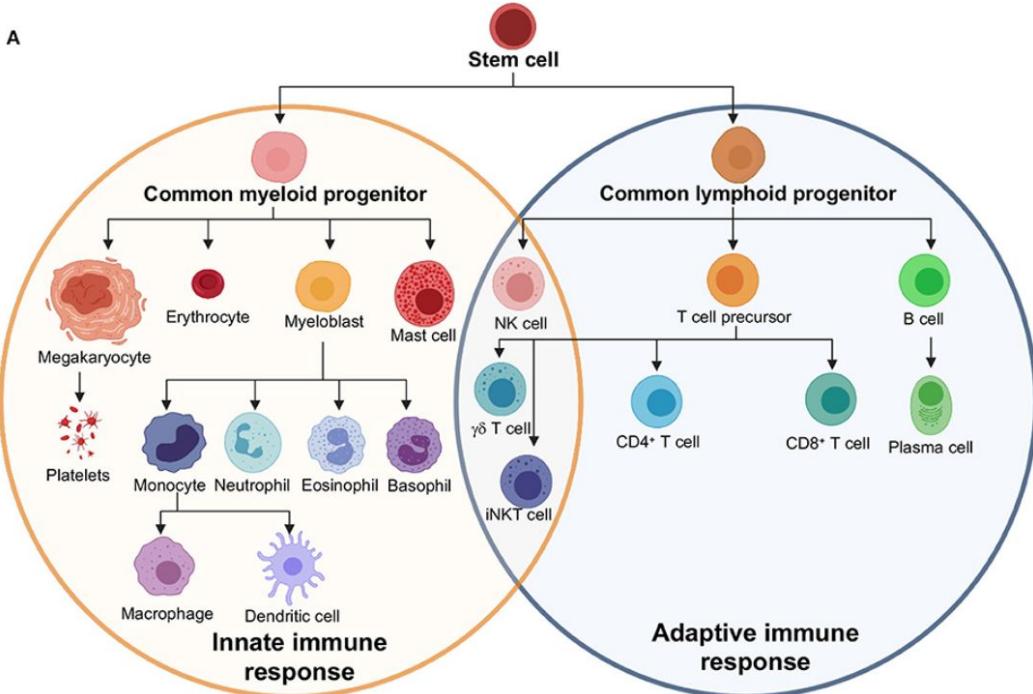
**Hematocrit:**

- Red blood cells



## Normal Blood:

♀ 37%–47% hematocrit  
 ♂ 42%–52% hematocrit

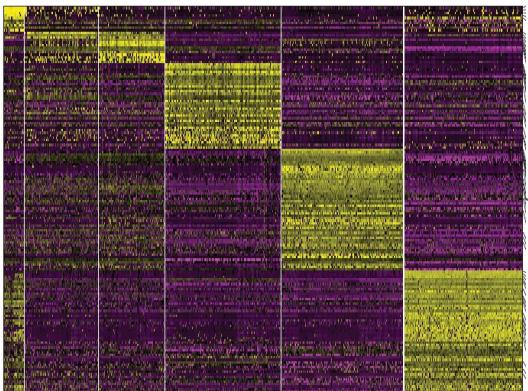


# What's in a drop of blood? Count the genes!



Sequencing

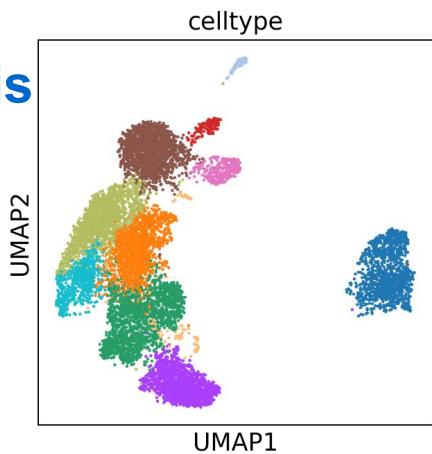
Genes



Cells

Low Expression  High Expression

Data analysis

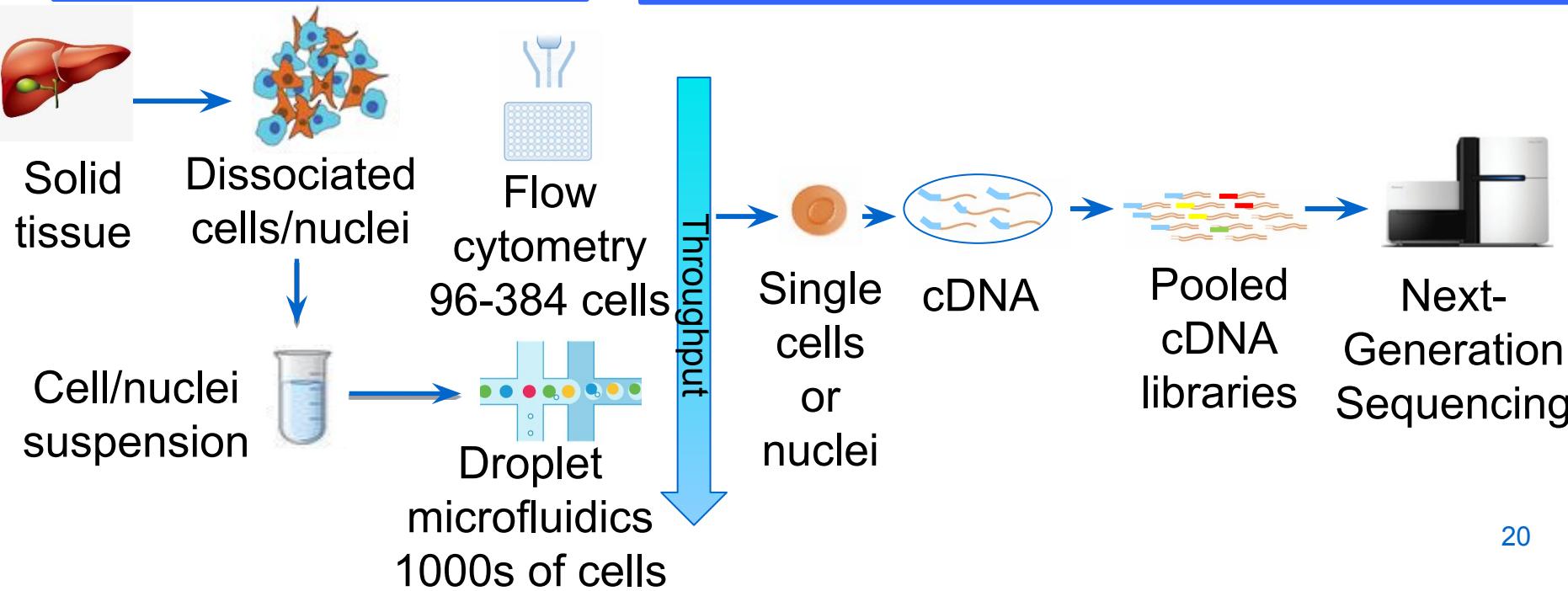


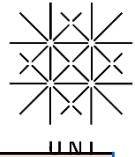
- B-cell
- CD4 T-cell
- CD8 T-cell
- DC
- NK cell
- monocyte CD14+
- monocyte CD16+
- naive CD4 T-cell
- naive CD8 T-cell
- pDC
- unknown

# Single-cell sequencing (scSeq) workflow

## Tissue dissociation

## Single cell capture and transcriptome sequencing





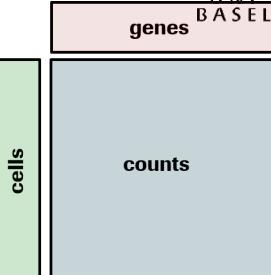
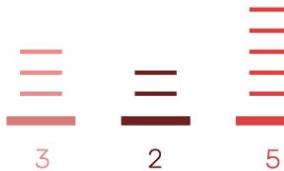
# A linearized workflow of scSeq data analysis

From short reads to gene-cell matrix

Alignment

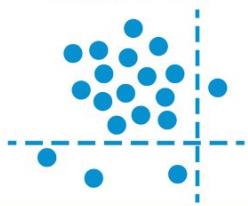


Quantification

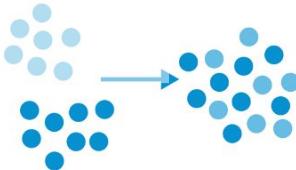


QC, filtering & normalization,  
dimensionality reduction, and  
clustering

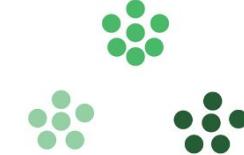
Quality control



Normalisation

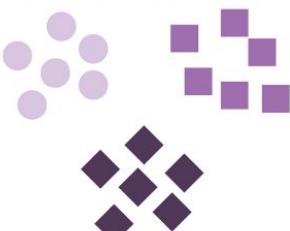


Clustering



Downstream analysis

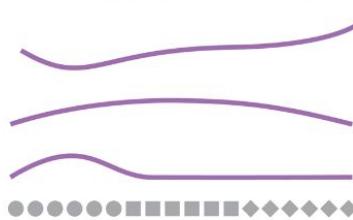
Differential expression



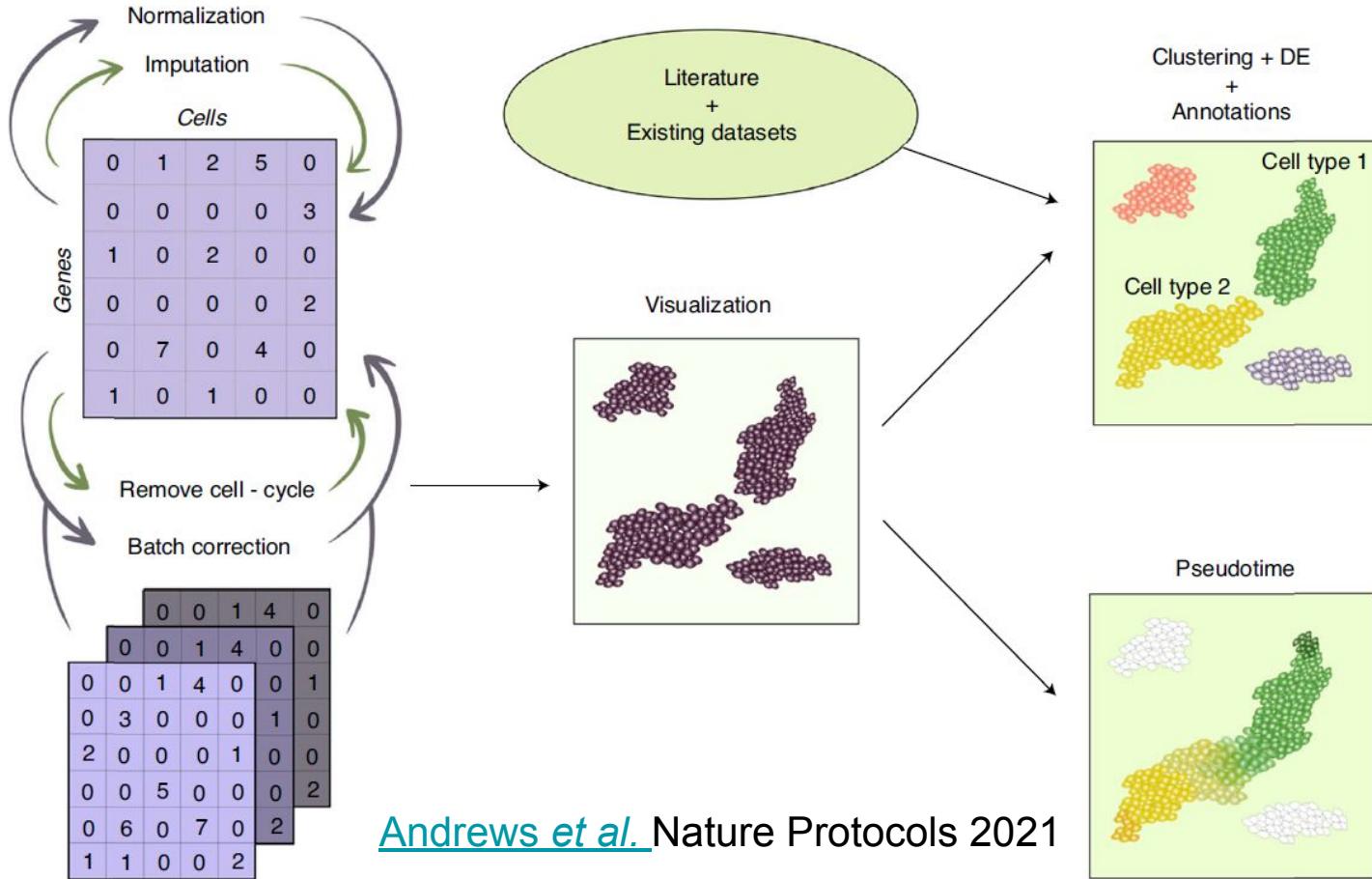
Marker genes



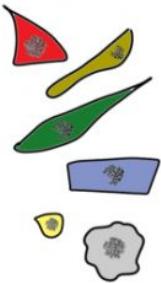
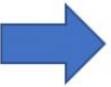
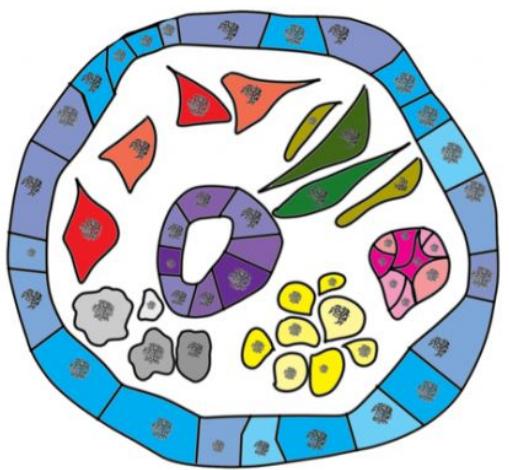
Expression patterns



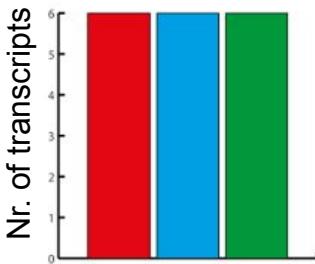
# Overview of the computational workflow



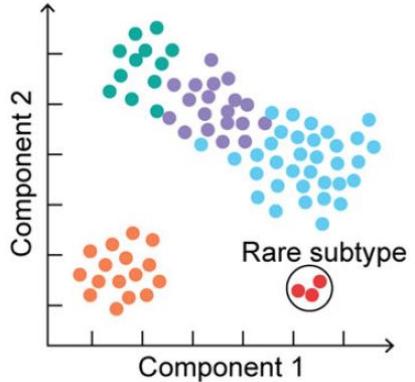
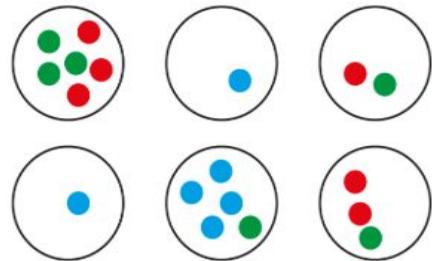
# Single-cell biology benefits both disease understanding and drug discovery



Bulk analysis

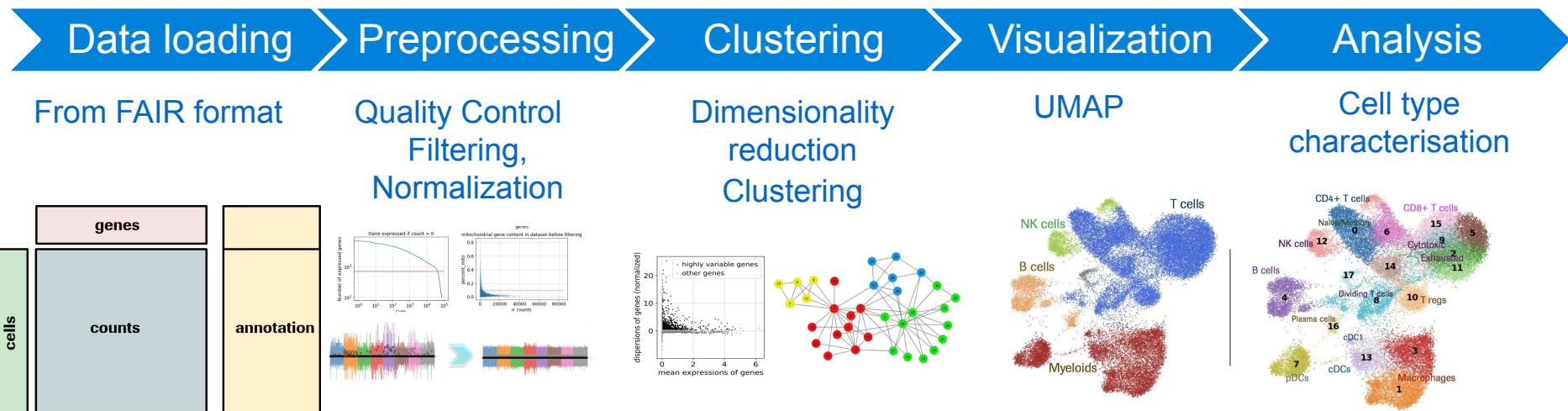


Single cell transcriptome analysis

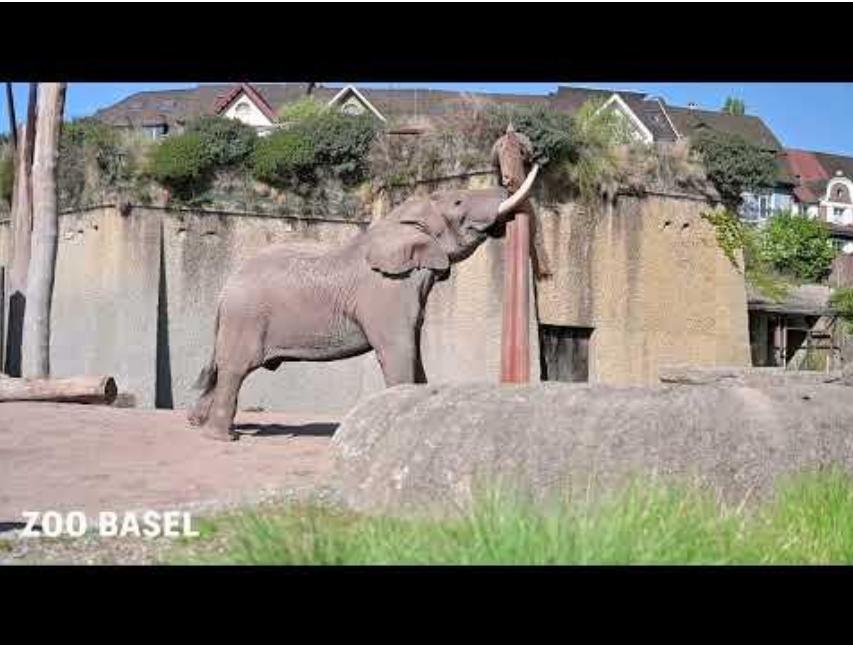


# BESCA: An open-source Python package for single-cell gene expression analysis

## An automated standard workflow

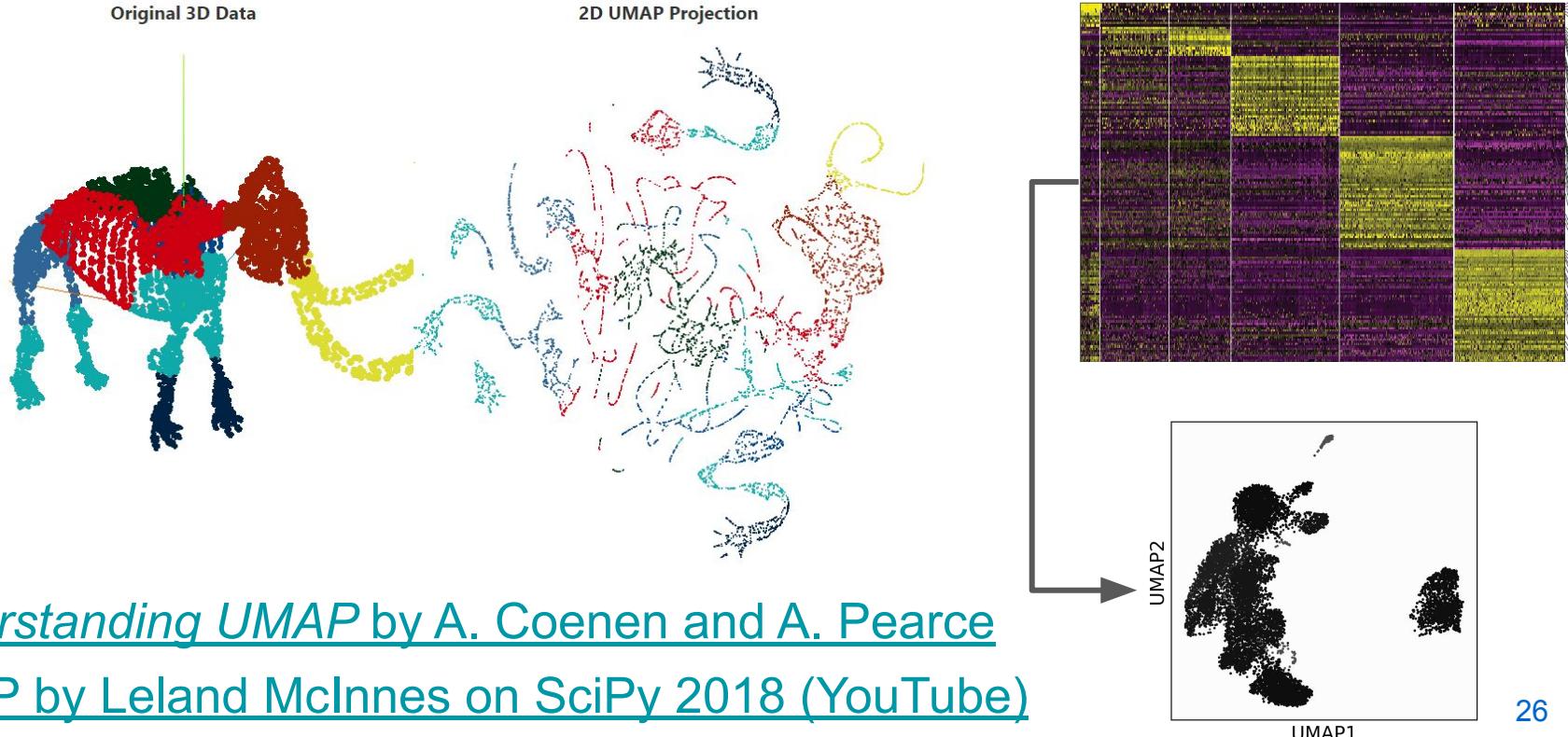


# How to represent voxels with pixels?

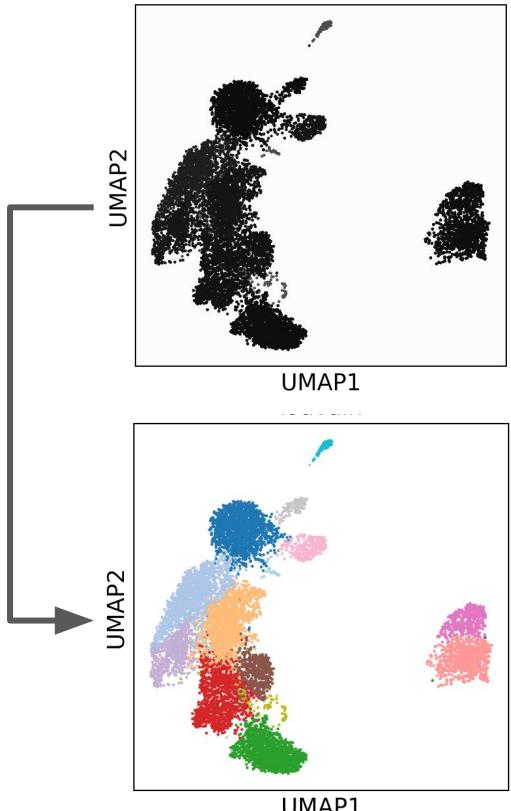
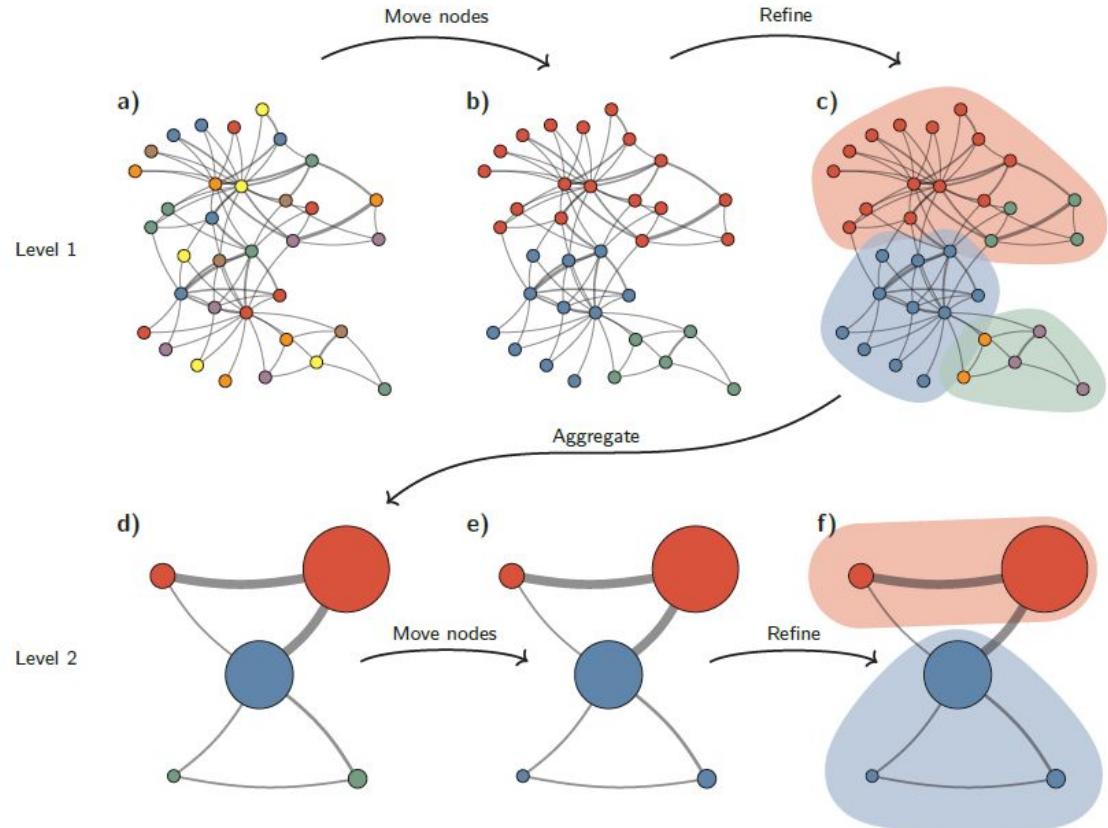


The elephant bull *Tusker* at Zolli Basel plays with a tree trunk on a post (2022)

# Uniform Manifold Approximation and Projection (UMAP) for dimension reduction

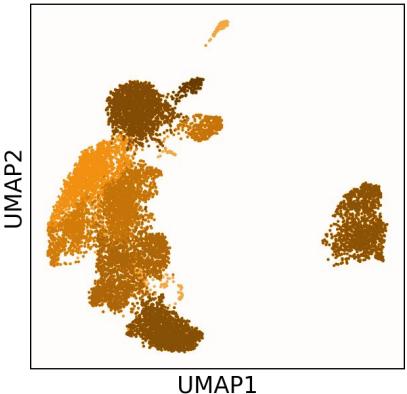


# The Leiden Algorithm for Community Detection

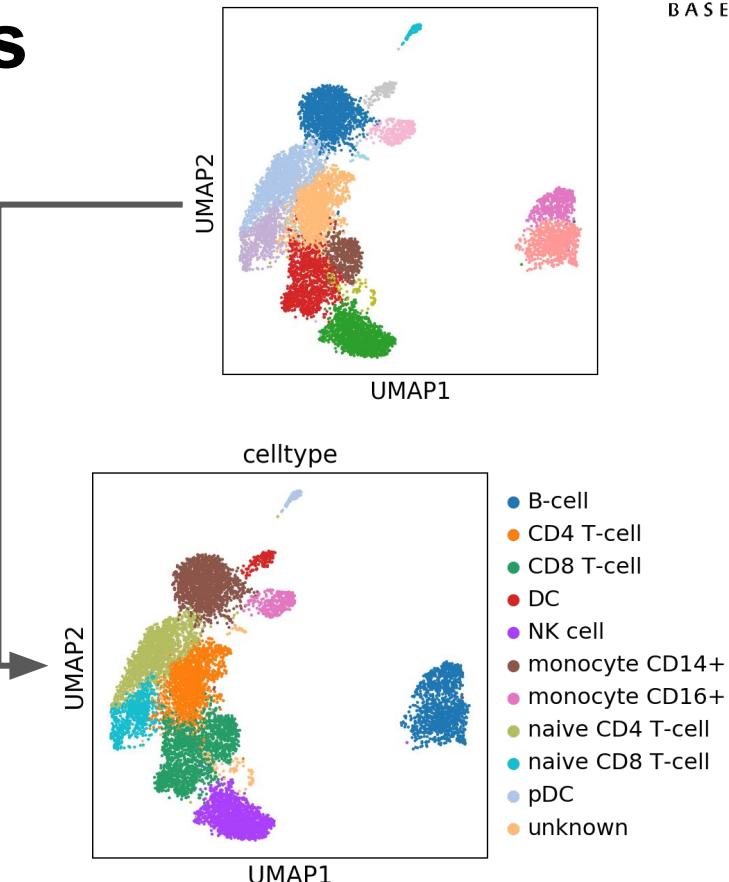


# Biological knowledge and visual inspection is used to annotate cell types

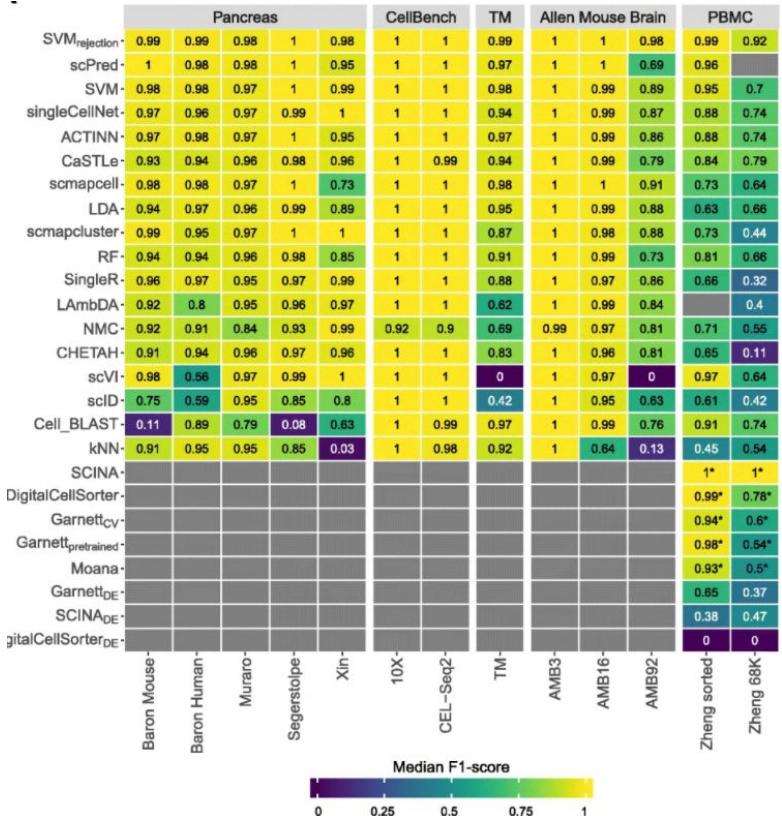
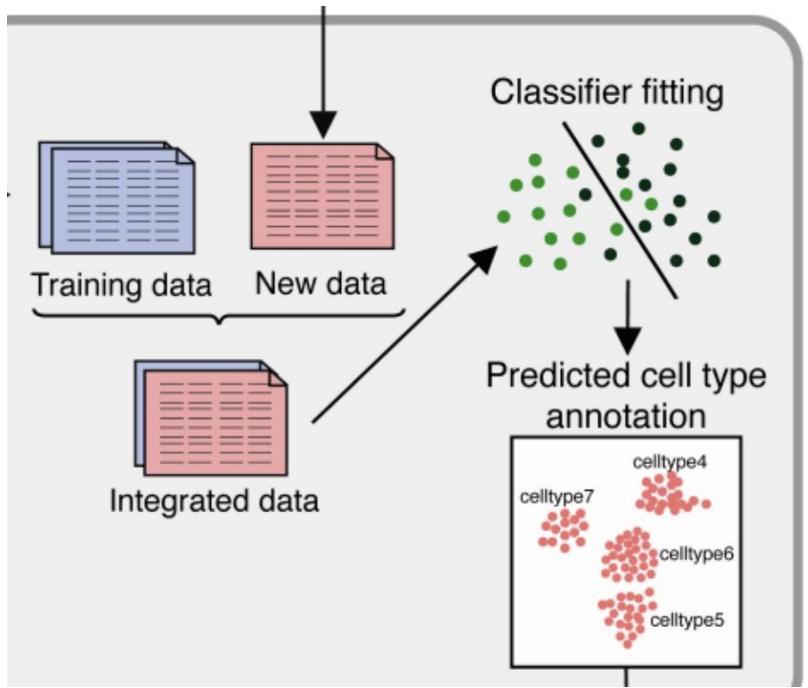
Heatmap  
of gene X



|            |        |        |       |       |        |        |       |        |
|------------|--------|--------|-------|-------|--------|--------|-------|--------|
| lymphocyte | PTPRC  |        |       |       |        |        |       |        |
| myeloid    | S100A8 | S100A9 | CST3  |       |        |        |       |        |
| Bcell      | CD19   | CD79A  | MS4A1 |       |        |        |       |        |
| Tcells     | CD3E   | CD3G   | CD3D  |       |        |        |       |        |
| CD4        | CD4    |        |       |       |        |        |       |        |
| CD8        | CD8A   | CD8B   |       |       |        |        |       |        |
| NKcell     | NKG7   | GNLY   | NCAM1 |       |        |        |       |        |
| monocyte   | CST3   | CSF1R  | ITGAM | CD14  | FCGR3A | FCGR3B |       |        |
| macrophage | CD14   | IL1B   | LYZ   | CD163 | ITGAX  | CD68   | CSF1R | FCGR3A |

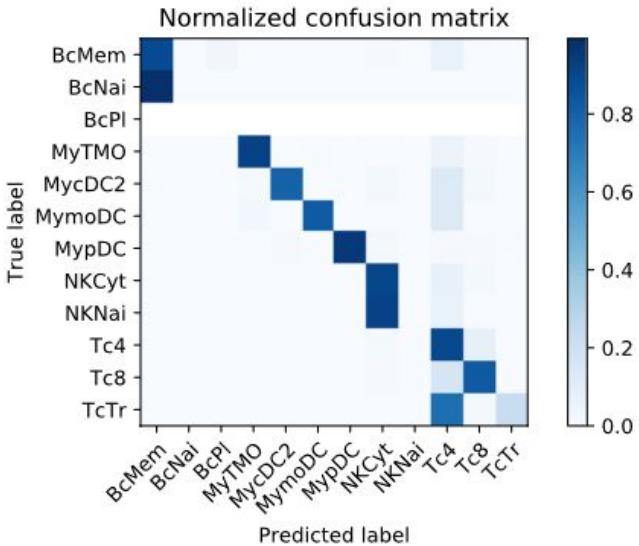
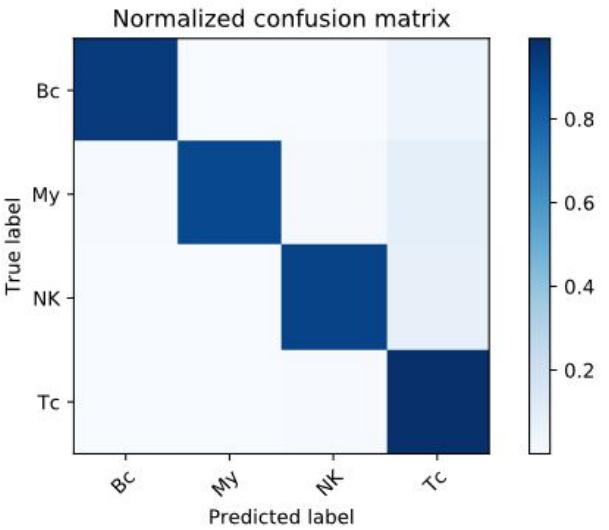
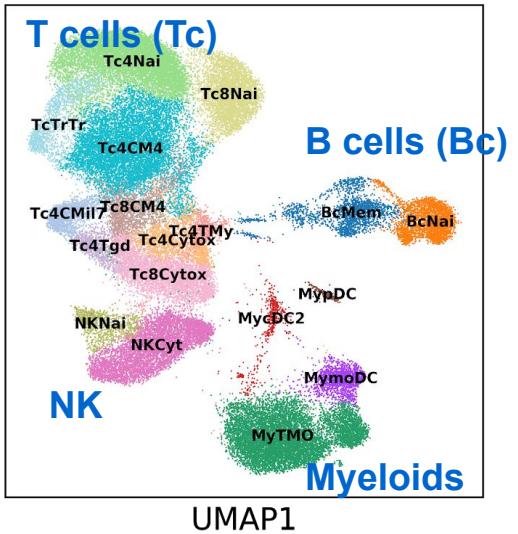


# Cell type annotation with machine learning



# A PBMC example of cell type annotation

UMAP2



- Broad level cell types, including B cells (Bc), Myeloid (My), NK cells (NK) and T cells (Tc), are successfully predicted.
- Missing and highly similar cell types cause challenges with increased granularity. Essential: reference data quality and knowledge of cell types.

# Single-cell biology is important in drug discovery

## Disease understanding:

disease-specific cell types  
and states



## Target identification:

expression pattern in  
health and disease across  
cell types



## Biomarker and patient stratification:

which genes should we measure  
in which cell type(s)?



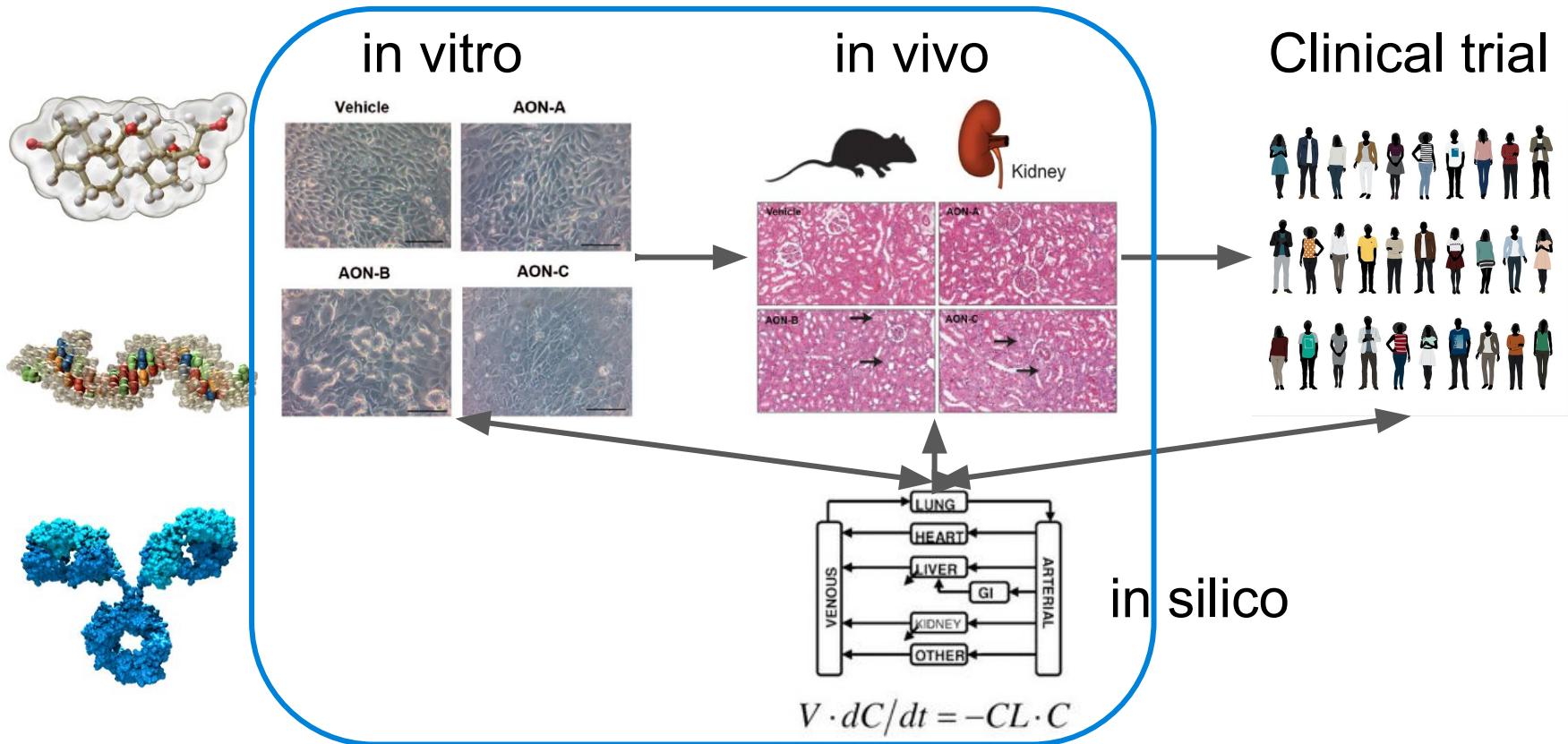
## MoA and safety

modelling: perturbation effect at single-cell level



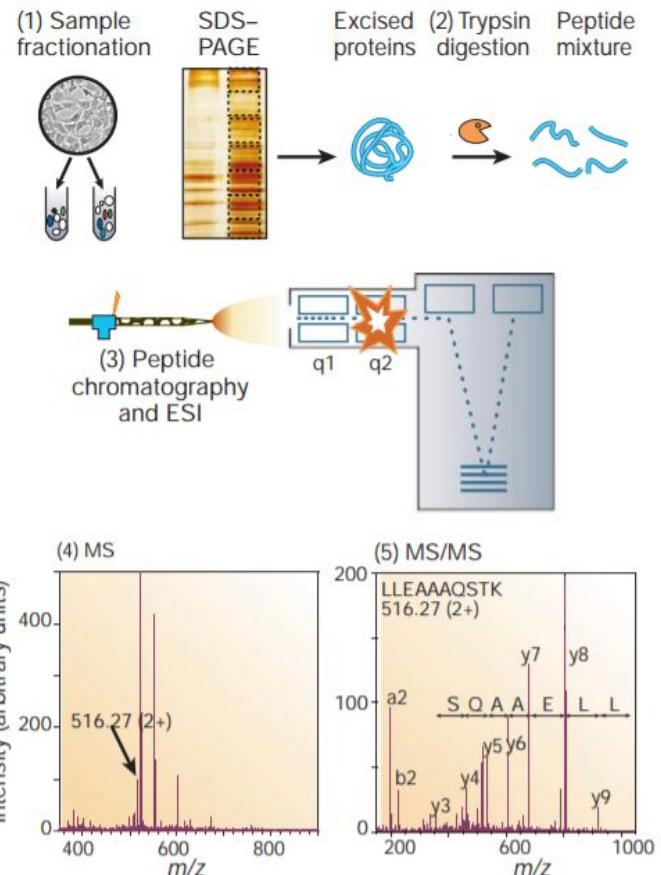
# End of Lecture 8

# Proteomics plays an important role in *in vitro/in vivo* translation



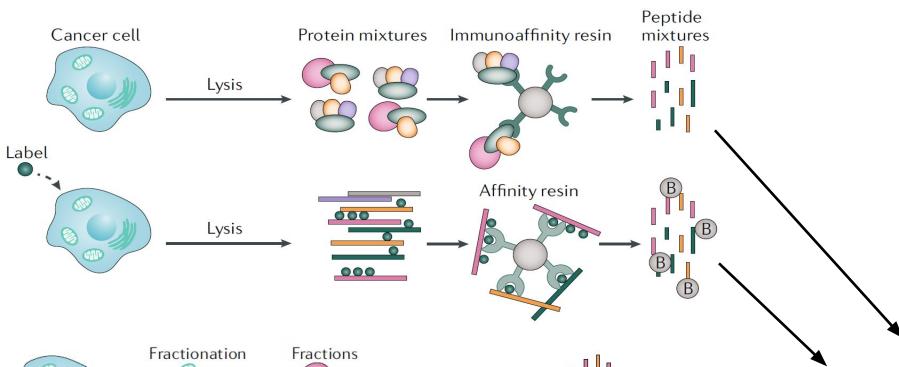
# Mass-spectrometry based proteomics

- **SDS-PAGE:** Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
- **ESI:** Electrospray ionization
- **q1/q2:** selection/collision/separation cells
- **MS:** Mass spectrometry
- **MS/MS:** tandem mass spectrometry



# Proteomics approaches for drug discovery

## Affinity purification



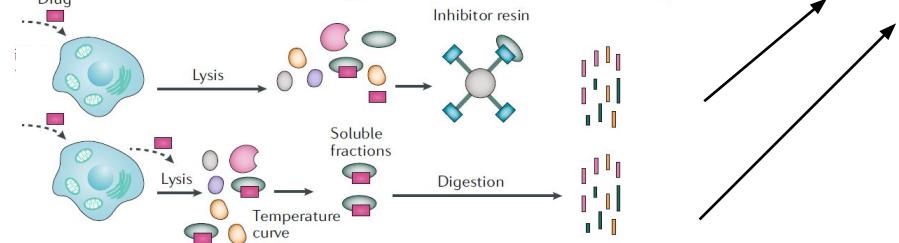
## Proximity labelling



## Organelle proteome profiling



## Post-translational modification (PTM) profiling



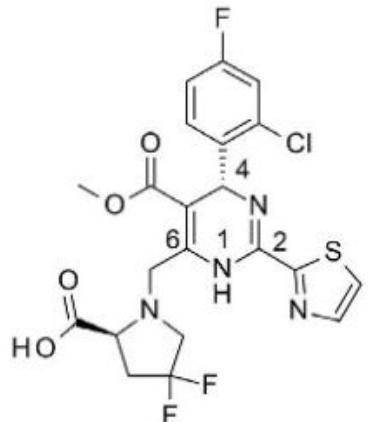
## Chemoaffinity enrichment

## Thermal proteome profiling

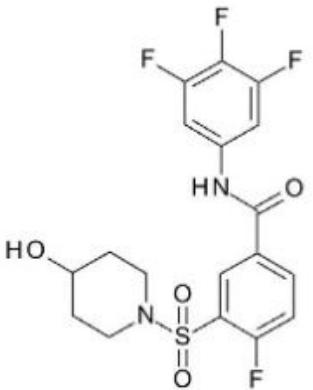
# Case 1: Differentiate two compounds that inhibit Hepatitis B Virus with similar mode of action

**a**

HAP\_R01

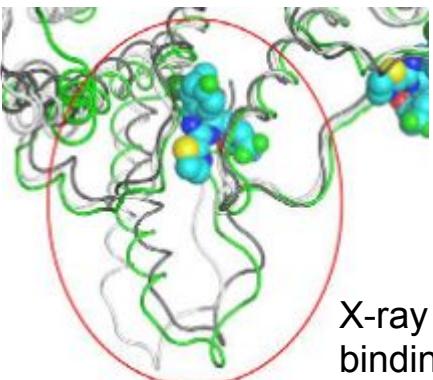


SBA\_R01

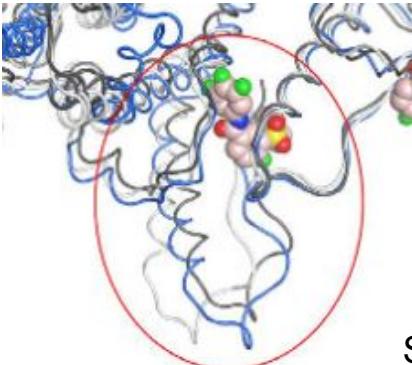


**b**

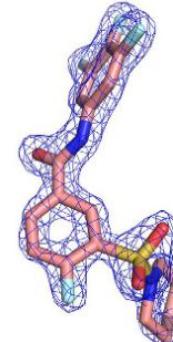
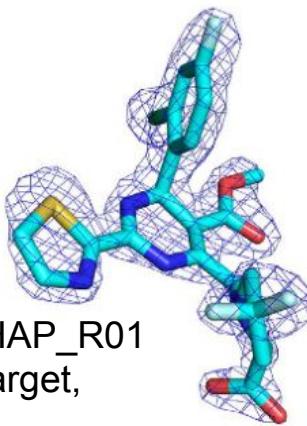
| Compound | IC <sub>50</sub> ( $\mu$ M) | HepG2.2.15 EC <sub>50</sub> ( $\mu$ M) | CC <sub>50</sub> ( $\mu$ M) |
|----------|-----------------------------|--|-----------------------------|
| HAP_R01  | 0.39 $\pm$ 0.13             | 0.0064 $\pm$ 0.0006                    | 34.8 $\pm$ 1.8              |
| SBA_R01  | 1.90 $\pm$ 0.22             | 0.26 $\pm$ 0.02                        | 8.05 $\pm$ 0.92             |



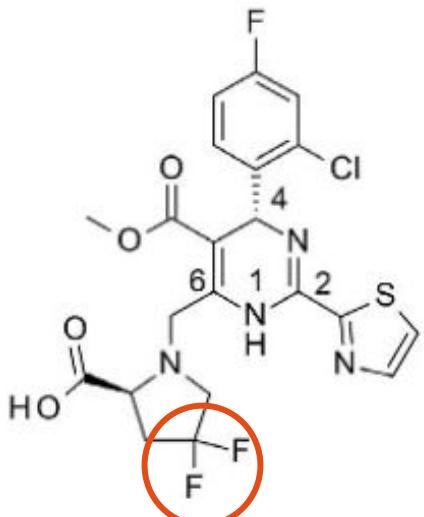
X-ray data of HAP\_R01 binding to its target, HBV capsid



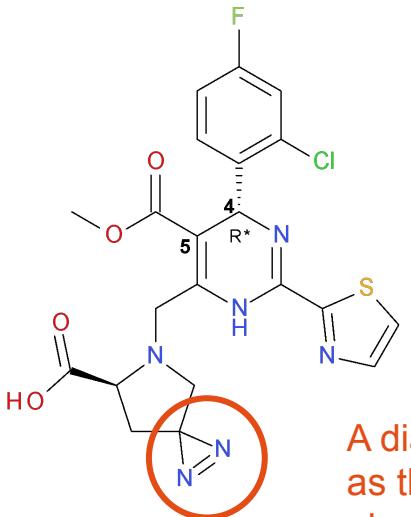
SBA\_R01



# Chemical probes: drug-like molecules to probe its mode of action

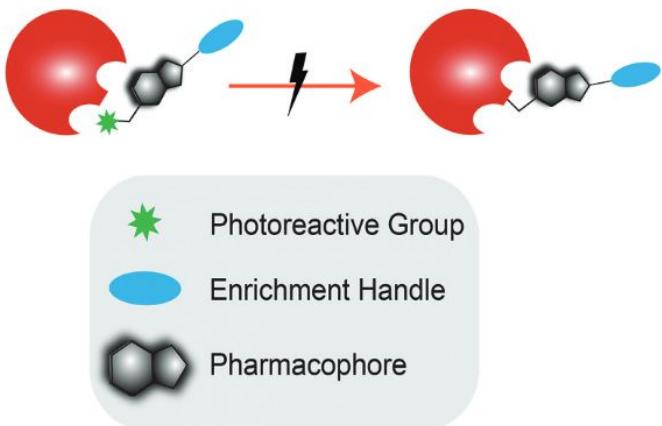


|         | $IC_{50}$ ( $\mu M$ ) |
|---------|-----------------------|
| HAP_R01 | $0.39 \pm 0.13$       |

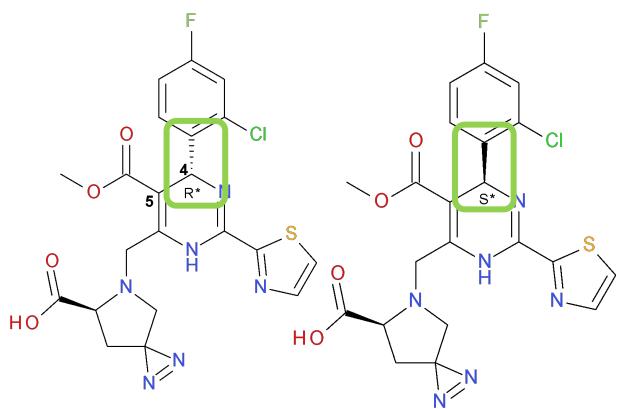


$EC_{50}$ : **0.040  $\mu M$**   
 $IC_{50}$ : **0.47  $\mu M$**

A diazirine group  
as the  
photoreactive  
group

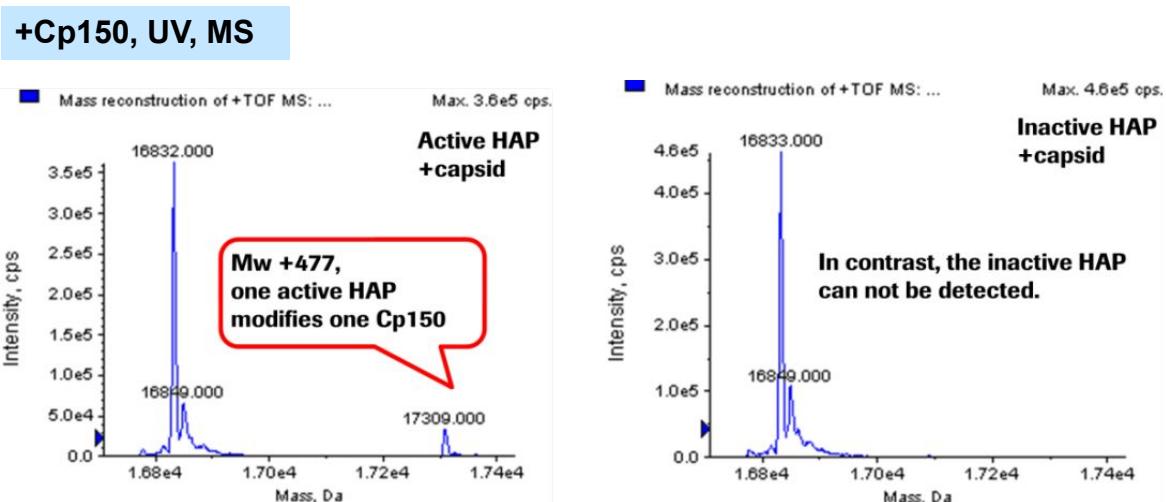


# Case 1 solved: Proteomics confirmed target binding and mapped the small molecule binding pocket



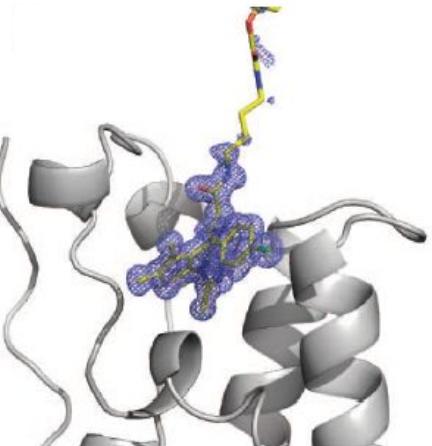
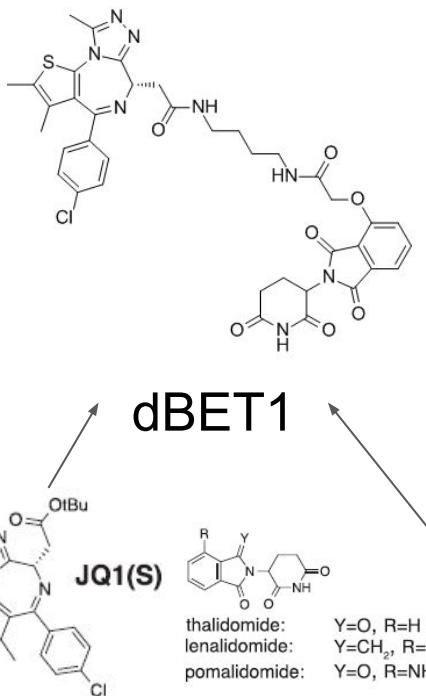
**RO-A**  
 $EC_{50}$ : 0.040  $\mu\text{M}$   
 $IC_{50}$ : 0.47  $\mu\text{M}$

**RO-B**  
 $EC_{50}$ : >1  $\mu\text{M}$   
 $IC_{50}$ : >100  $\mu\text{M}$

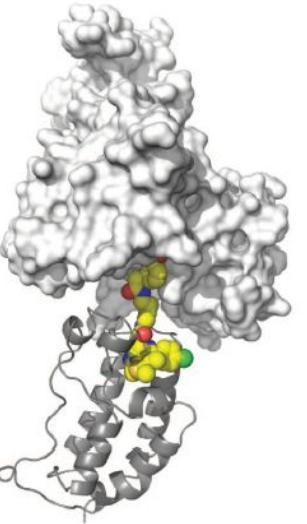


Proteolytic digestion/LC-MS/MS identified labelling site **Y118 (Y=Tyrosine)** of HBV capsid protein. More photoaffinity probes identified labelling sites at **R127 (R=Arginine)** and **Y38**.

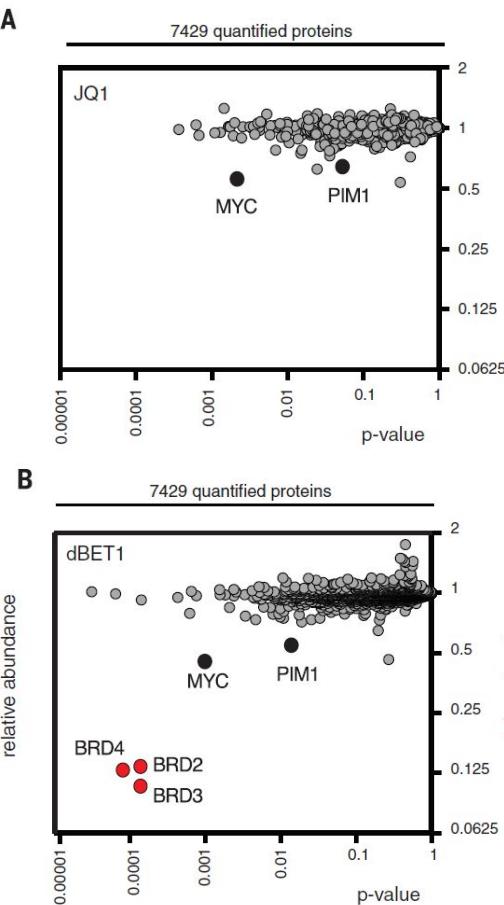
# Case 2: Confirmation of selective degradation of protein target *in vivo*



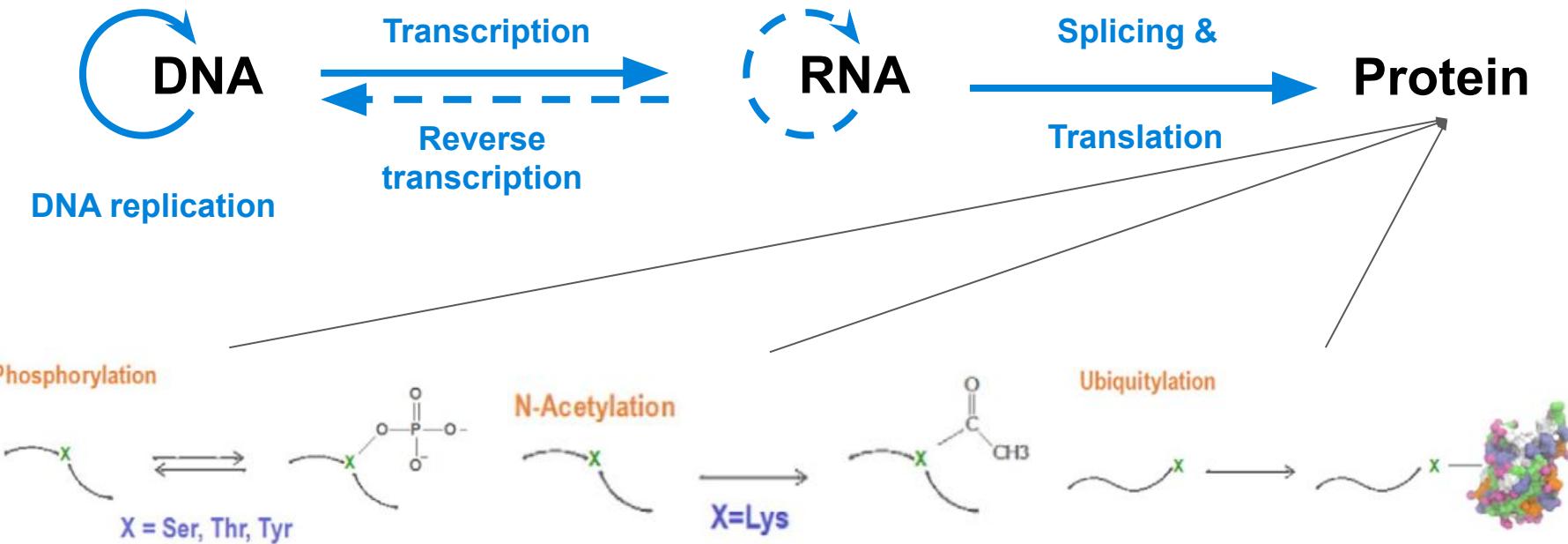
Crystal structure of dBET1 binding to its target BRD4



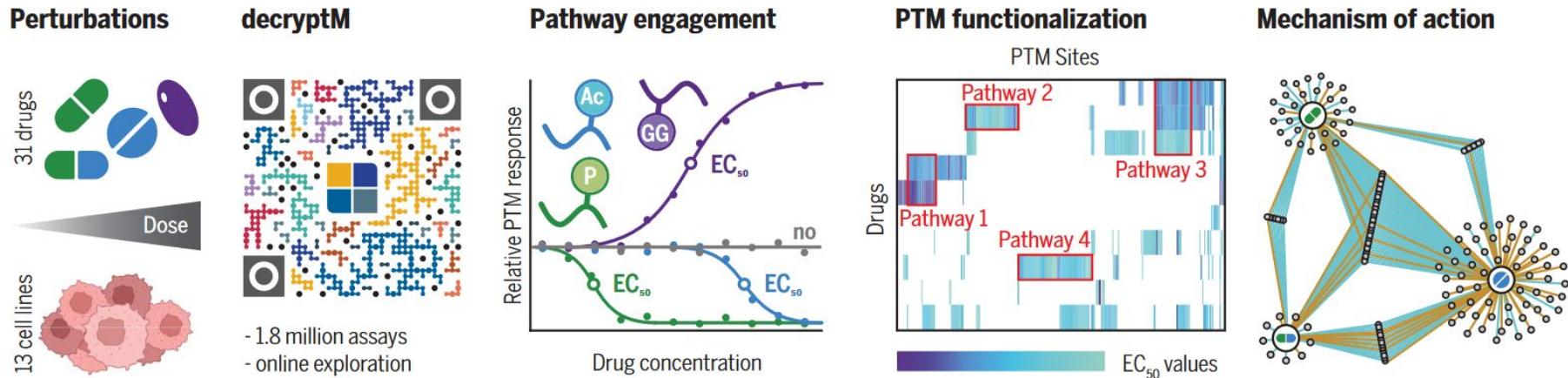
Docking of dBET1-BRD4 to DDB1-CRBN structure



# Protein post-translational modifications (PTMs) offer an additional layer of regulation

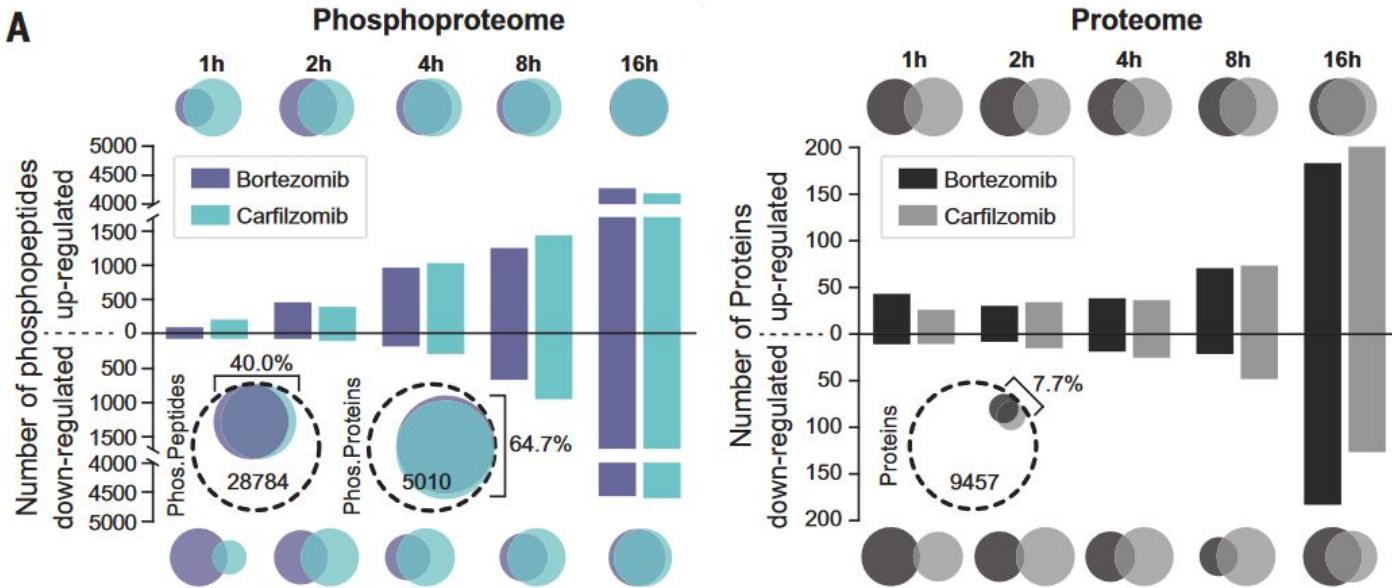


# Case 3: Millions of PTM profiles induced by drugs in cancer cell lines



**decryptM (Nature 2023):** Following the dose-dependent treatment of cancer cells with drugs, quantitative mass spectrometry records dose-response of thousands of posttranslationally modified peptides. EC<sub>50</sub>: half-maximal effective concentration; Ac, acetylation; GG, ubiquitinylation; P, phosphorylation.

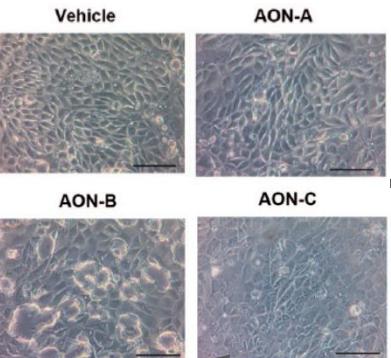
# PTM and proteomics characterize MoA of drugs



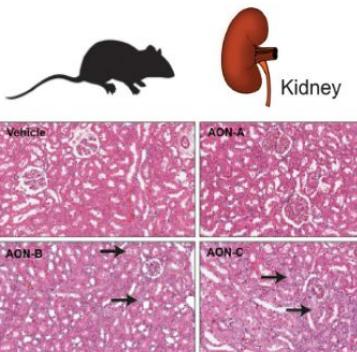
Bortezomib (BTZ) and carfilzomib (CFZ) both treat multiple myeloma by inhibiting the proteasome by reversible covalent (BTZ) or irreversible (CFZ) binding to the protease PSMB5. Time-series data show both the dynamics and the converging signaling.

# Dose prediction based on pharmacology and toxicology before entry into human

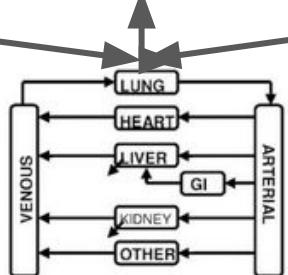
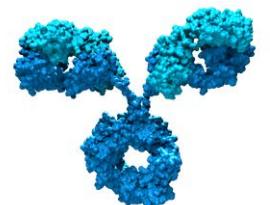
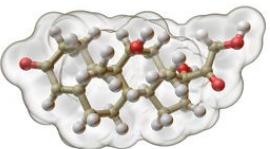
**in vitro**



**in vivo**



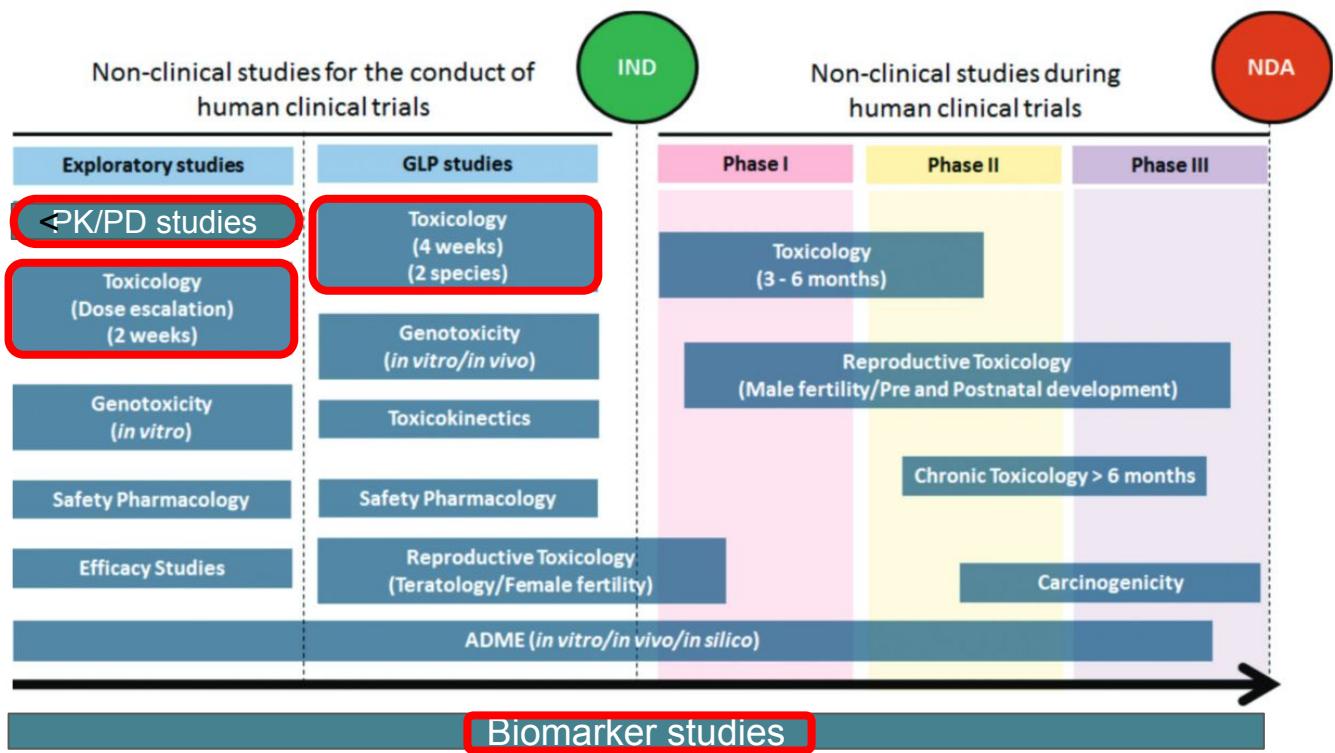
**Clinical trial**



**in silico**

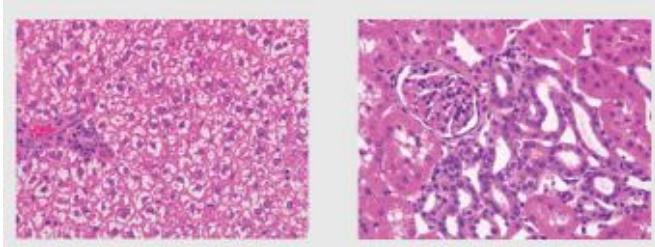
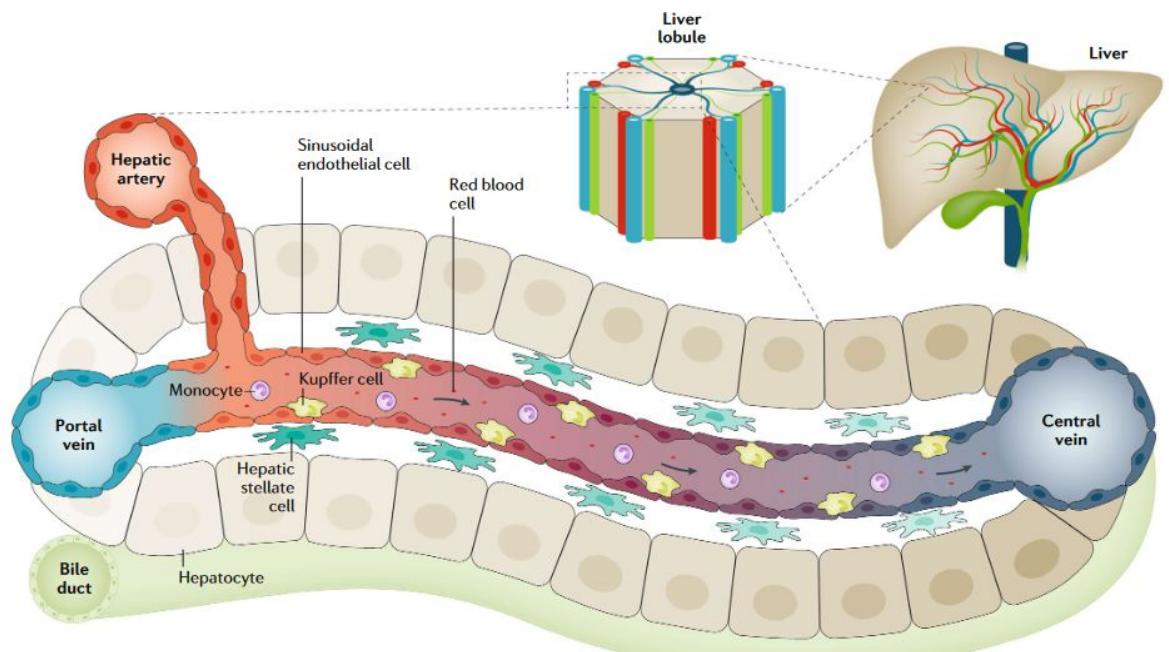
$$V \cdot dC/dt = -CL \cdot C$$

# Current practices of non-clinical studies in drug development

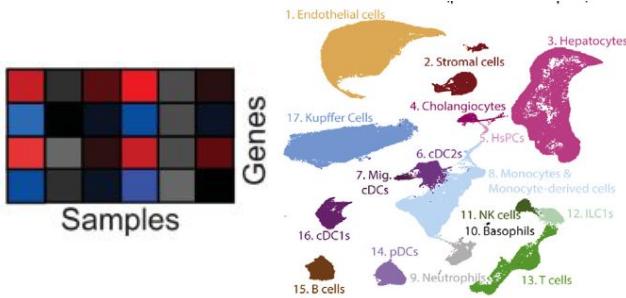


- IND: Investigational New Drug application
- NDA: New Drug Application
- GLP: Good Lab Practice
- Red boxes: Focus areas of this and coming lectures

# Current practices of profiling and understanding toxicology: an example with liver



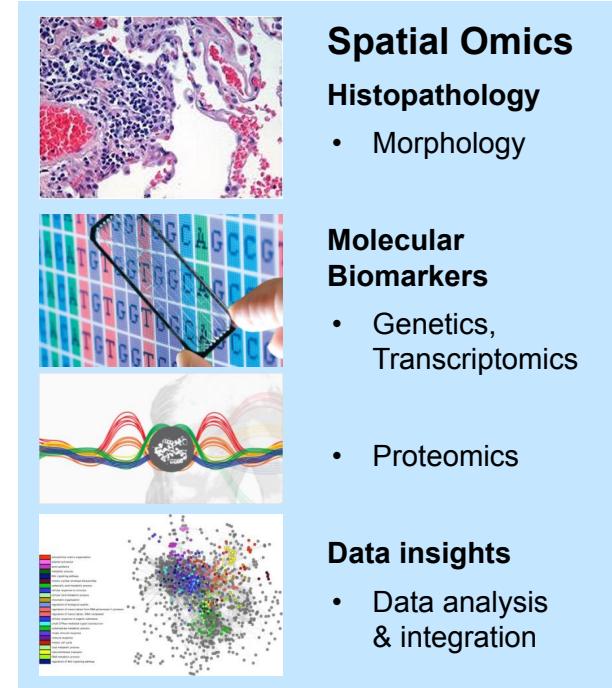
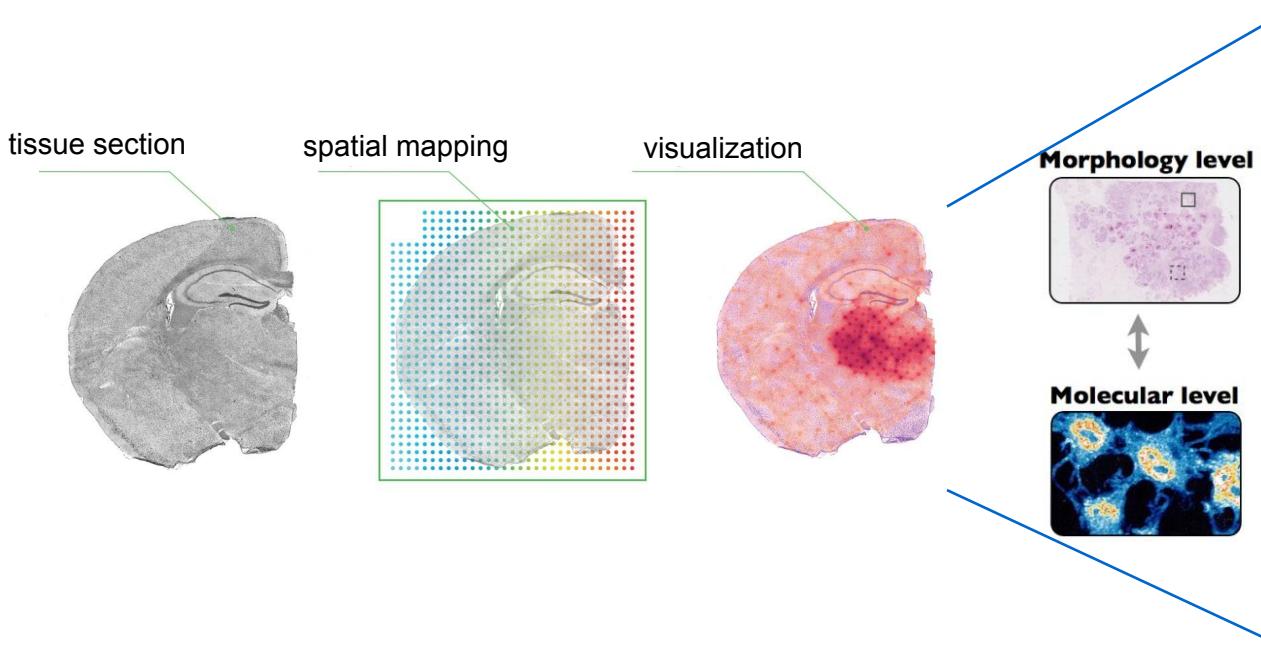
## Histopathology



## Omics

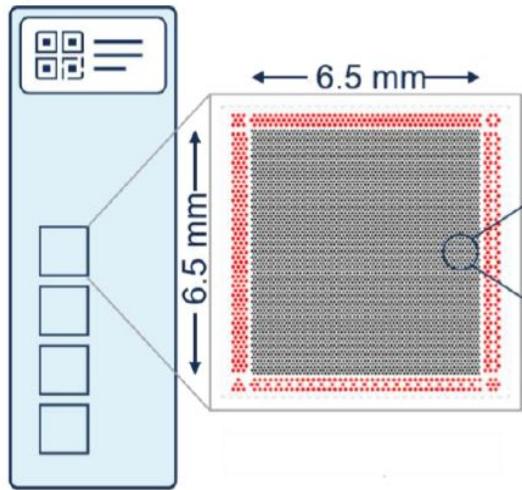
[Liver structure and anatomy \(YouTube Video\)](#)

# Spatially resolved omics complement histopathology

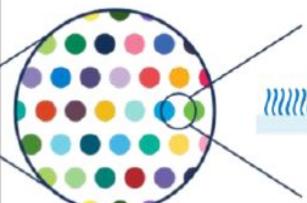


# An example: 10x VISIUM Technology

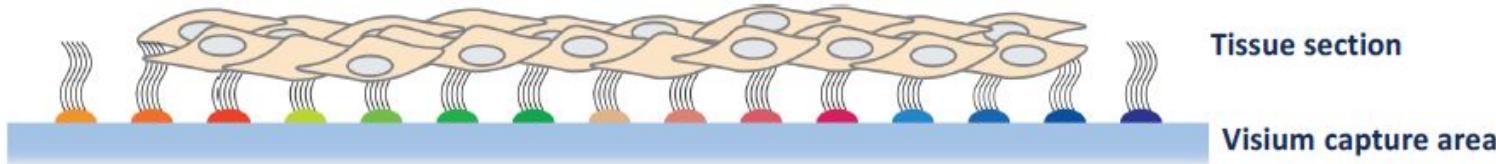
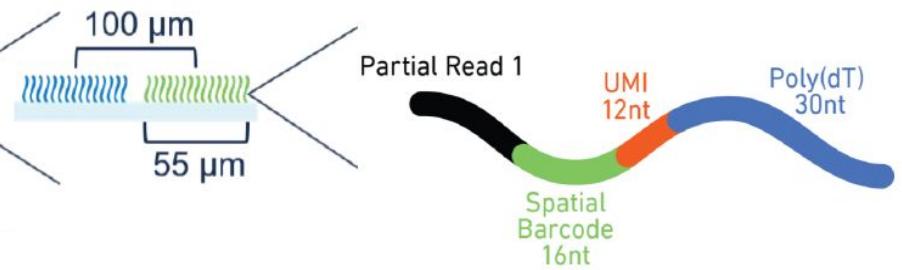
Visium Spatial Gene Expression Slide



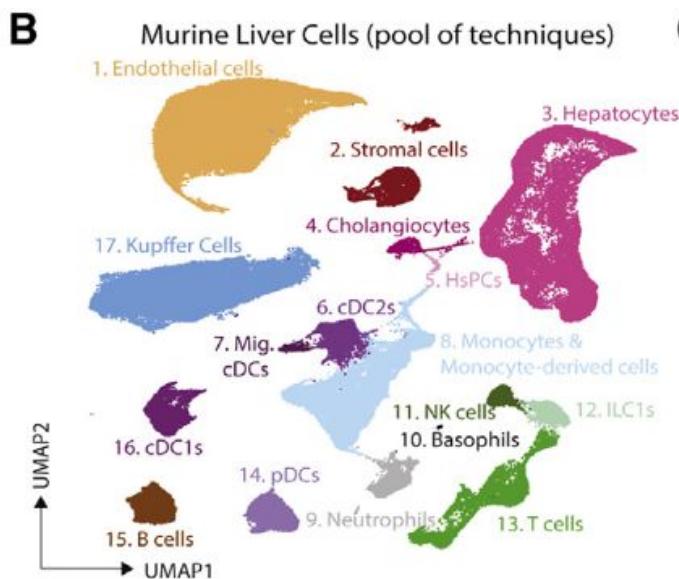
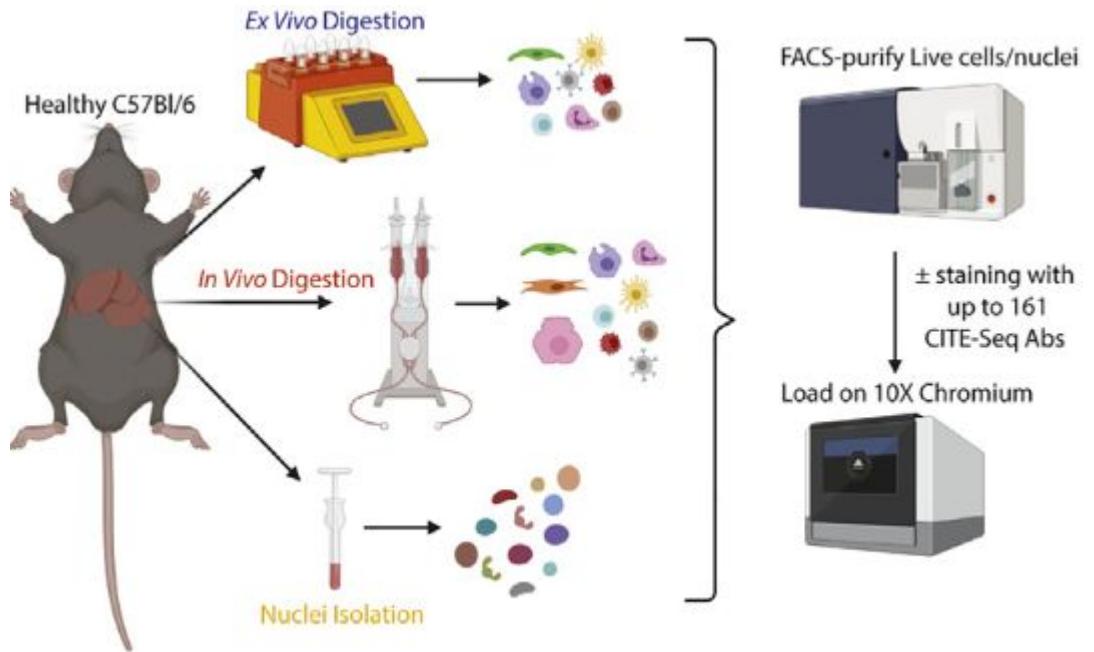
Capture Area with ~5000 Barcoded Spots



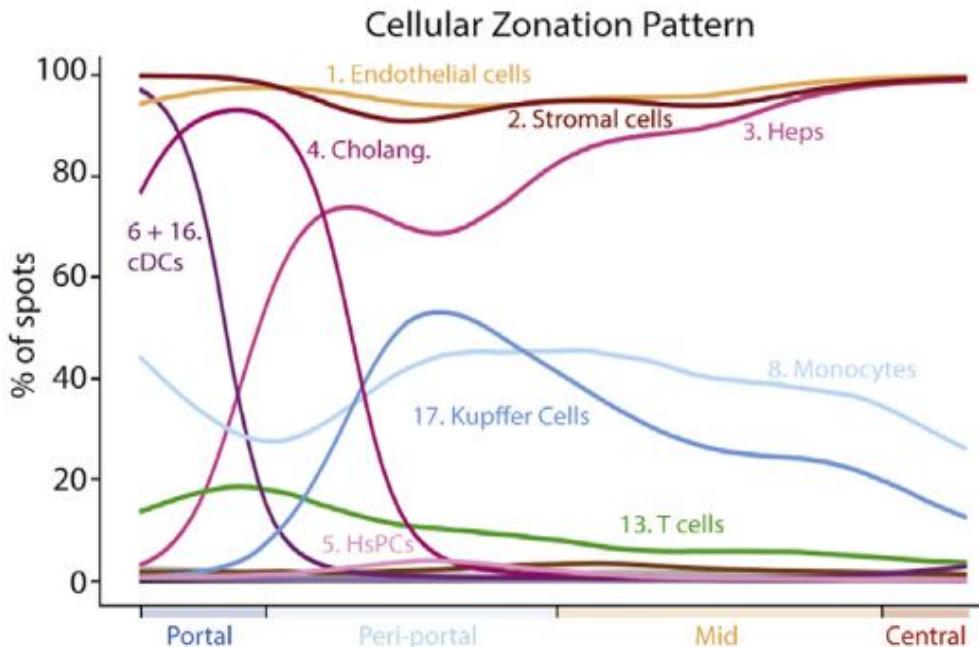
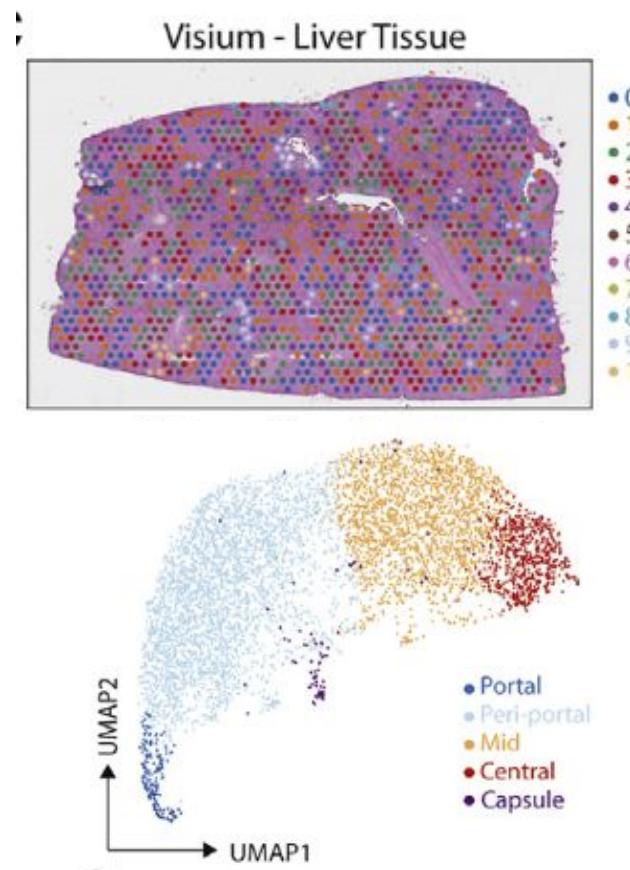
Visium Gene Expression Barcoded Spots



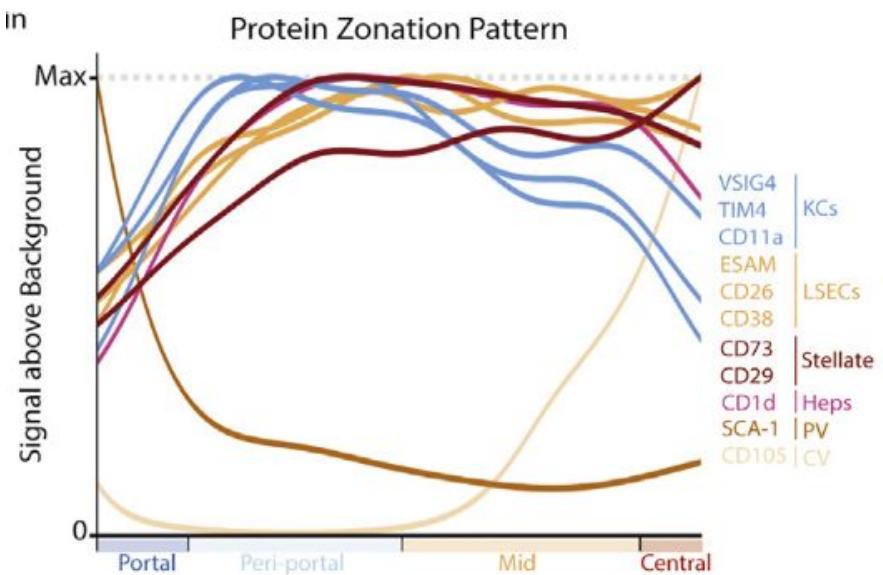
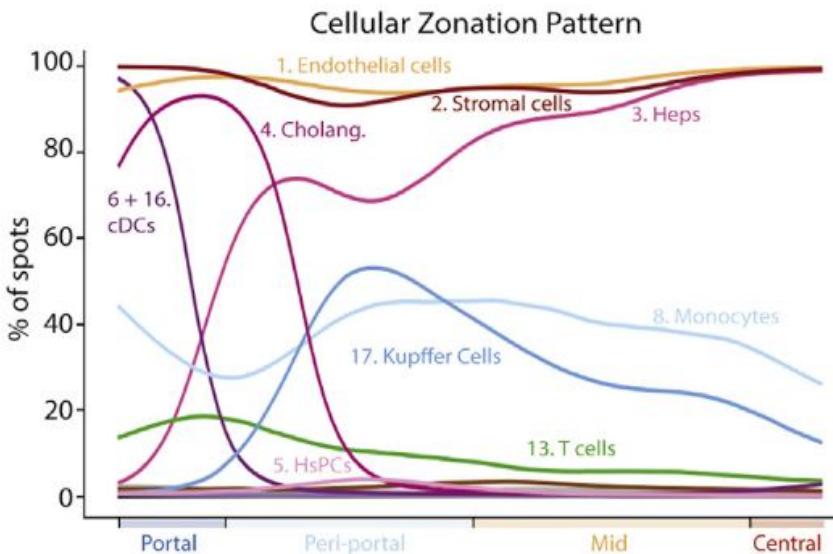
# Spatial and single-cell expression of liver cells



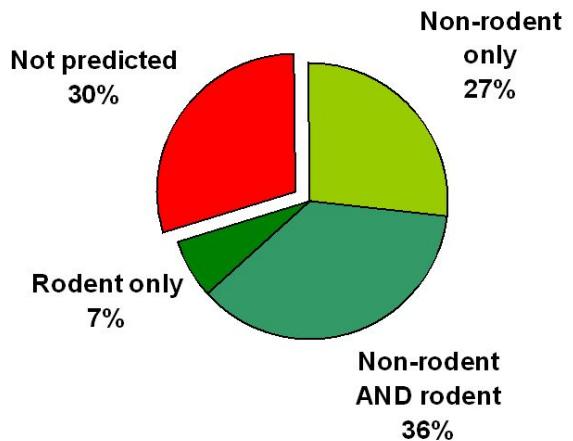
# Spatial and single-cell expression of liver cells



# Spatial mRNA and protein expression data empowers digital pathology and biological understanding



# How predictive is animal safety testing for humans? It depends on modality and therapeutic classes.



[Regul Toxicol Pharmacol. 2000;32:56-67](https://doi.org/10.1016/j.reprotox.2009.08.001)

| Target organ of ADRs      | Small molecule drugs |                  | Large molecule drugs |                  |
|---------------------------|----------------------|------------------|----------------------|------------------|
|                           | % of ADRs            | % of correlation | % of ADRs            | % of correlation |
| Gastrointestinal          | 21                   | 80               | 14                   | 19               |
| Neurological              | 20                   | 34               | 11                   | 4                |
| Hepatobiliary             | 11                   | 73               | 8                    | 21               |
| Hematological             | 8                    | 75               | 8                    | 80               |
| Cutaneous                 | 5                    | 56               | 9                    | 22               |
| Systemic                  | 5                    | 45               | 8                    | 20               |
| Cardiovascular            | 4                    | 61               | 6                    | 0                |
| Ocular                    | 5                    | 64               | 5                    | 83               |
| Musculoskeletal           | 3                    | 16               | 5                    | 0                |
| Metabolic                 | 4                    | 50               | 3                    | 43               |
| Faucal/oral               | 4                    | 41               | 3                    | 38               |
| Urinary                   | 3                    | 61               | 3                    | 14               |
| Respiratory               | 1                    | 45               | 5                    | 32               |
| Infection                 | 0.4                  | 100              | 6                    | 68               |
| Nasal                     | 1                    | 27               | 2                    | 33               |
| Application site reaction | 1                    | 100              | 3                    | 81               |
| Others                    | 3                    | 45               | 1                    | 80               |

# Conclusions

- We predict efficacy and safety profiles of drugs by studying the mechanism and mode of action (MoA).
- Bulk and single-cell RNA sequencing, and proteomics based on mass spectrometry (MS) are essential tools for understanding MoA of drug candidates.
- Spatial omics combines imaging and omics technologies to offer spatially resolved data of biological systems. Their use in animal models and human samples has the potential to improve translational studies.

# Offline activities of Module IV (optional)

Perform your own single-cell data analysis to get first-hand experience working with high-dimensional biological data.

- If you are new to the topic, please use [the PBMC tutorial of Scanpy \(python\)](#) or [the PBMC tutorial of Seurat \(R\)](#).
- If you have experience with such data already, checkout [the NBIS workshop on single-cell sequencing data analysis](#) to cover advanced topics such as spatial transcriptomics and trajectory inference.

# References

1. Figures: [Lumen Learning](#), [Exploring Nature](#), [National Geographic](#), [Platelet cells](#) (Graham Beards, CC-BY-SA 4.0), [Lymphocytes](#) (Nicolas Grandjean, CC-BY-SA 3.0), [Adipocytes](#) (Public Domain), [Hepatocytes](#) (CC-BY-NC 2.0), [Neurons and Glia](#) (Public Domain), [Blood](#) (CC 3.0), [Blood Cells](#) (By A. Rad and M. Häggström. CC-BY-SA 3.0 license), [A selective JAK3 inhibitor](#) (London Lab/Weizmann institute)
2. Sender, Ron, Shai Fuchs, and Ron Milo. 2016. "Revised Estimates for the Number of Human and Bacteria Cells in the Body." PLoS Biology 14 (8). <https://doi.org/10.1371/journal.pbio.1002533>.
3. [www.evocell-itn.eu](http://www.evocell-itn.eu);
4. Macaulay, Iain C., and Thierry Voet. 2014. "Single Cell Genomics: Advances and Future Perspectives." PLOS Genetics 10 (1): e1004126. <https://doi.org/10.1371/journal.pgen.1004126>.
5. Pryor, Rosina, Povilas Norvaisas, Georgios Marinos, Lena Best, Louise B. Thingholm, Leonor M. Quintaneiro, Wouter De Haes, et al. 2019. "Host-Microbe-Drug-Nutrient Screen Identifies Bacterial Effectors of Metformin Therapy." Cell 178 (6): 1299-1312.e29. <https://doi.org/10.1016/j.cell.2019.08.003>.
6. Cully, Megan. 2019. "Microbiome Therapeutics Go Small Molecule." *Nature Reviews Drug Discovery* 18 (July): 569. <https://doi.org/10.1038/d41573-019-00122-8>.
7. Duscha, Alexander, Barbara Gisevius, Sarah Hirschberg, Nissan Yissachar, Gabriele I. Stangl, Eva Eilers, Verian Bader, et al. 2020. "Propionic Acid Shapes the Multiple Sclerosis Disease Course by an Immunomodulatory Mechanism." Cell 180 (6): 1067-1080.e16. <https://doi.org/10.1016/j.cell.2020.02.035>.
8. Pryor, Rosina, Povilas Norvaisas, Georgios Marinos, Lena Best, Louise B. Thingholm, Leonor M. Quintaneiro, Wouter De Haes, et al. 2019. "Host-Microbe-Drug-Nutrient Screen Identifies Bacterial Effectors of Metformin Therapy." Cell 178 (6): 1299-1312.e29. <https://doi.org/10.1016/j.cell.2019.08.003>.
9. Zimmermann, Michael, Maria Zimmermann-Kogadeeva, Rebekka Wegmann, and Andrew L. Goodman. 2019. "Mapping Human Microbiome Drug Metabolism by Gut Bacteria and Their Genes." Nature 570 (7762): 462. <https://doi.org/10.1038/s41586-019-1291-3>.
10. Shin, Hyun Kil, Young-Mook Kang, and Kyoung Tai No. 2016. "Predicting ADME Properties of Chemicals." In *Handbook of Computational Chemistry*, edited by Jerzy Leszczynski, 1–37. Dordrecht: Springer Netherlands. [https://doi.org/10.1007/978-94-007-6169-8\\_59-1](https://doi.org/10.1007/978-94-007-6169-8_59-1).

# References (continued)

11. Mädler, Sophia Clara, Alice Julien-Laferriere, Luis Wyss, Miroslav Phan, Albert S. W. Kang, Eric Ulrich, Roland Schmucki, et al. 2020. "Besca, a Single-Cell Transcriptomics Analysis Toolkit to Accelerate Translational Research." *BioRxiv*, September, 2020.08.11.245795. <https://doi.org/10.1101/2020.08.11.245795>.
12. Andrews, Tallulah S., Vladimir Yu Kiselev, Davis McCarthy, and Martin Hemberg. 2021. "Tutorial: Guidelines for the Computational Analysis of Single-Cell RNA Sequencing Data." *Nature Protocols* 16 (1): 1–9. <https://doi.org/10.1038/s41596-020-00409-w>.
13. Sturm, Gregor, Francesca Finotello, Florent Petitprez, Jitao David Zhang, Jan Baumbach, Wolf H. Fridman, Markus List, and Tatsiana Aneichyk. 2019. "Comprehensive Evaluation of Transcriptome-Based Cell-Type Quantification Methods for Immuno-Oncology." *Bioinformatics* 35 (14): i436–45. <https://doi.org/10.1093/bioinformatics/btz363>.
14. Villani, Alexandra-Chloé, Rahul Satija, Gary Reynolds, Siranush Sarkizova, Karthik Shekhar, James Fletcher, Morgane Griesbeck, et al. 2017. "Single-Cell RNA-Seq Reveals New Types of Human Blood Dendritic Cells, Monocytes, and Progenitors." *Science* 356 (6335): eaah4573. <https://doi.org/10.1126/science.aah4573>.
15. Finotello, Francesca, Clemens Mayer, Christina Plattner, Gerhard Laschober, Dietmar Rieder, Hubert Hackl, Anne Krogsdam, et al. 2019. "Molecular and Pharmacological Modulators of the Tumor Immune Contexture Revealed by Deconvolution of RNA-Seq Data." *Genome Medicine* 11 (1): 34. <https://doi.org/10.1186/s13073-019-0638-6>.
16. Fridman, Wolf H., Laurence Zitvogel, Catherine Sautès–Fridman, and Guido Kroemer. 2017. "The Immune Contexture in Cancer Prognosis and Treatment." *Nature Reviews Clinical Oncology* 14 (12): 717–34. <https://doi.org/10.1038/nrclinonc.2017.101>.
17. Moisan, Annie, Marcel Gubler, Jitao David Zhang, Yann Tessier, Kamille Dumong Erichsen, Sabine Sewing, Régine Gérard, et al. 2017. "Inhibition of EGF Uptake by Nephrotoxic Antisense Drugs In Vitro and Implications for Preclinical Safety Profiling." *Molecular Therapy - Nucleic Acids* 6 (March): 89–105. <https://doi.org/10.1016/j.omtn.2016.11.006>.
18. Chang, Chia-Yu, Hsiao-Chien Ting, Ching-Ann Liu, Hong-Lin Su, Tzyy-Wen Chiou, Horng-Jyh Harn, and Shinn-Zong Lin. 2018. "Induced Pluripotent Stem Cells: A Powerful Neurodegenerative Disease Modeling Tool for Mechanism Study and Drug Discovery." *Cell Transplantation* 27 (June): 096368971877540. <https://doi.org/10.1177/0963689718775406>.

# References (continued)

19. Takahashi, Toshio. 2019. "Organoids for Drug Discovery and Personalized Medicine." *Annual Review of Pharmacology and Toxicology* 59 (1): 447–62. <https://doi.org/10.1146/annurev-pharmtox-010818-021108>.
20. Budayeva, Hanna G., and Donald S. Kirkpatrick. 2020. "Monitoring Protein Communities and Their Responses to Therapeutics." *Nature Reviews Drug Discovery* 19 (6): 414–26. <https://doi.org/10.1038/s41573-020-0063-y>.
21. Lukonin, Ilya, Denise Serra, Ludivine Challet Meylan, Katrin Volkmann, Janine Baaten, Rui Zhao, Shelly Meeusen, et al. 2020. "Phenotypic Landscape of Intestinal Organoid Regeneration." *Nature* 586 (7828): 275–80. <https://doi.org/10.1038/s41586-020-2776-9>.
22. Drawnel, Faye M., Stefano Boccardo, Michael Prummer, Frédéric Delobel, Alexandra Graff, Michael Weber, Régine Gérard, et al. 2014. "Disease Modeling and Phenotypic Drug Screening for Diabetic Cardiomyopathy Using Human Induced Pluripotent Stem Cells." *Cell Reports* 9 (3): 810–20. <https://doi.org/10.1016/j.celrep.2014.09.055>.
23. Traag, Vincent, Ludo Waltman, and Nees Jan van Eck. 2019. "From Louvain to Leiden: Guaranteeing Well-Connected Communities." *Scientific Reports* 9 (1): 5233. <https://doi.org/10.1038/s41598-019-41695-z>.
24. *Understanding UMAP*, Andy Coenen and Adam Pearce, <https://pair-code.github.io/understanding-umap/>
25. How exactly UMAP works, Nikolay Oskolkov, <https://towardsdatascience.com/how-exactly-umap-works-13e3040e1668>
26. McInnes, Leland, and John Healy. 2018. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." ArXiv:1802.03426 [Cs, Stat], February. <http://arxiv.org/abs/1802.03426>.
27. Zappia, Luke, Belinda Phipson, and Alicia Oshlack. 2018. "Exploring the Single-Cell RNA-Seq Analysis Landscape with the ScRNA-Tools Database." *PLOS Computational Biology* 14 (6): e1006245. <https://doi.org/10.1371/journal.pcbi.1006245>.
28. Abdelaal, Tamim, Lieke Michielsen, Davy Cats, Dylan Hoogduin, Hailiang Mei, Marcel J. T. Reinders, and Ahmed Mahfouz. 2019. "A Comparison of Automatic Cell Identification Methods for Single-Cell RNA Sequencing Data." *Genome Biology* 20 (1): 194. <https://doi.org/10.1186/s13059-019-1795-z>.
29. Janas, Maja M., Mark K. Schlegel, Carole E. Harbison, Vedat O. Yilmaz, Yongfeng Jiang, Rubina Parmar, Ivan Zlatev, et al. 2018. "Selection of GalNAc-Conjugated SiRNAs with Limited off-Target-Driven Rat Hepatotoxicity." *Nature Communications* 9 (1): 723. <https://doi.org/10.1038/s41467-018-02989-4>.

# References (continued)

30. Jackson, Aimee L., and Peter S. Linsley. 2010. "Recognizing and Avoiding SiRNA Off-Target Effects for Target Identification and Therapeutic Application." *Nature Reviews Drug Discovery* 9 (1): 57–67. <https://doi.org/10.1038/nrd3010>.
31. Romond, Edward H., Edith A. Perez, John Bryant, Vera J. Suman, Charles E. Geyer, Nancy E. Davidson, Elizabeth Tan-Chiu, et al. 2005. "Trastuzumab plus Adjuvant Chemotherapy for Operable HER2-Positive Breast Cancer." *New England Journal of Medicine* 353 (16): 1673–84. <https://doi.org/10.1056/NEJMoa052122>.
32. Gao, Jinxu, Adelphe Mfuh, Yuka Amako, and Christina M. Woo. 2018. "Small Molecule Interactome Mapping by Photoaffinity Labeling Reveals Binding Site Hotspots for the NSAIDs." *Journal of the American Chemical Society* 140 (12): 4259–68. <https://doi.org/10.1021/jacs.7b11639>.
33. Bollag, Gideon, James Tsai, Jiazhong Zhang, Chao Zhang, Prabha Ibrahim, Keith Nolop, and Peter Hirth. 2012. "Vemurafenib: The First Drug Approved for BRAF -Mutant Cancer." *Nature Reviews Drug Discovery* 11 (11): 873–86. <https://doi.org/10.1038/nrd3847>.
34. Luebker, Stephen A., and Scott A. Koepsell. 2019. "Diverse Mechanisms of BRAF Inhibitor Resistance in Melanoma Identified in Clinical and Preclinical Studies." *Frontiers in Oncology* 9. <https://doi.org/10.3389/fonc.2019.00268>.
35. Kimball's Biology Page, <http://www.biology-pages.info/>
36. Molina, Daniel Martinez, Rozbeh Jafari, Marina Ignatushchenko, Takahiro Seki, E. Andreas Larsson, Chen Dan, Lekshmy Sreekumar, Yihai Cao, and Pär Nordlund. 2013. "Monitoring Drug Target Engagement in Cells and Tissues Using the Cellular Thermal Shift Assay." *Science* 341 (6141): 84–87. <https://doi.org/10.1126/science.1233606>.
37. Zhou, Zheng, Taishan Hu, Xue Zhou, Steffen Wildum, Fernando Garcia-Alcalde, Zhiheng Xu, Daitze Wu, et al. 2017. "Heteroaryldihydropyrimidine (HAP) and Sulfamoylbenzamide (SBA) Inhibit Hepatitis B Virus Replication by Different Molecular Mechanisms." *Scientific Reports* 7 (1): 42374. <https://doi.org/10.1038/srep42374>.
38. Dai, Jiang, Yi-Jiao Huang, Xinhua He, Ming Zhao, Xinzhen Wang, Zhao-Shan Liu, Wen Xue, et al. 2019. "Acetylation Blocks CGAS Activity and Inhibits Self-DNA-Induced Autoimmunity." *Cell* 176 (6): 1447–1460.E14. <https://doi.org/10.1016/j.cell.2019.01.016>.

# References (continued)

39. Hart, Charles P. 2005. "Finding the Target after Screening the Phenotype." *Drug Discovery Today* 10 (7): 513–19.  
[https://doi.org/10.1016/S1359-6446\(05\)03415-X](https://doi.org/10.1016/S1359-6446(05)03415-X).
40. Ziegler, Slava, Sonja Sievers, and Herbert Waldmann. 2021. "Morphological Profiling of Small Molecules." *Cell Chemical Biology* 28 (3): 300–319.  
<https://doi.org/10.1016/j.chembiol.2021.02.012>.
41. Winter, Georg E., Dennis L. Buckley, Joshiawa Paulk, Justin M. Roberts, Amanda Souza, Sirano Dhe-Paganon, and James E. Bradner. 2015. "Phthalimide Conjugation as a Strategy for in Vivo Target Protein Degradation." *Science* 348 (6241): 1376–81.  
<https://doi.org/10.1126/science.aab1433>.
42. Aebersold, Ruedi, and Matthias Mann. 2016. "Mass-Spectrometric Exploration of Proteome Structure and Function." *Nature* 537 (7620): 347–55.  
<https://doi.org/10.1038/nature19949>.
43. Zhou, Jing C., Bob Feller, Bill Hinsberg, Geeta Sethi, Paul Feldstein, Joshua Hihath, Erkin Seker, Maria Marco, Andre Knoesen, and Robert Miller. 2015. "Immobilization-Mediated Reduction in Melting Temperatures of DNA–DNA and DNA–RNA Hybrids: Immobilized DNA Probe Hybridization Studied by SPR." *Colloids and Surfaces A: Physicochemical and Engineering Aspects* 481 (September): 72–79.  
<https://doi.org/10.1016/j.colsurfa.2015.04.046>.
44. Hagedorn, Peter H., Malene Pontoppidan, Tina S. Bisgaard, Marco Berrera, Andreas Dieckmann, Martin Ebeling, Marianne R. Møller, et al. 2018. "Identifying and Avoiding Off-Target Effects of RNase H-Dependent Antisense Oligonucleotides in Mice." *Nucleic Acids Research* 46 (11): 5366–80.  
<https://doi.org/10.1093/nar/gky397>.
45. Rehmsmeier, Marc, Peter Steffen, Matthias Hochsmann, and Robert Giegerich. 2004. "Fast and Effective Prediction of MicroRNA/Target Duplexes." *RNA* (New York, N.Y.) 10 (10): 1507–17. <https://doi.org/10.1261/rna.5248604>.
46. Tyanova, Stefka, Tikira Temu, and Juergen Cox. 2016. "The MaxQuant Computational Platform for Mass Spectrometry-Based Shotgun Proteomics." *Nature Protocols* 11 (12): 2301–19. <https://doi.org/10.1038/nprot.2016.136>.
47. xkcd: <https://xkcd.com/1217/>

# References (continued)

48. Murin, Charles D. "Considerations of Antibody Geometric Constraints on NK Cell Antibody Dependent Cellular Cytotoxicity." *Frontiers in Immunology* 11 (2020). <https://www.frontiersin.org/article/10.3389/fimmu.2020.01635>.
49. Marx, Vivien. "Method of the Year: Spatially Resolved Transcriptomics." *Nature Methods* 18, no. 1 (January 2021): 9–14. <https://doi.org/10.1038/s41592-020-01033-y>.
50. Andrade, E. L., A. F. Bento, J. Cavalli, S. K. Oliveira, R. C. Schwanke, J. M. Siqueira, C. S. Freitas, R. Marcon, and J. B. Calixto. "Non-Clinical Studies in the Process of New Drug Development - Part II: Good Laboratory Practice, Metabolism, Pharmacokinetics, Safety and Dose Translation to Clinical Studies." *Brazilian Journal of Medical and Biological Research* 49 (December 12, 2016). <https://doi.org/10.1590/1414-431X20165646>.
51. Olson, H., G. Betton, D. Robinson, K. Thomas, A. Monro, G. Kolaja, P. Lilly, et al. "Concordance of the Toxicity of Pharmaceuticals in Humans and in Animals." *Regulatory Toxicology and Pharmacology: RTP* 32, no. 1 (August 2000): 56–67. <https://doi.org/10.1006/rtpb.2000.1399>.
52. Tamaki, Chihiro, Takashi Nagayama, Masamichi Hashiba, Masato Fujiyoshi, Masanori Hizue, Hiroshi Kodaira, Minoru Nishida, et al. "Potentials and Limitations of Nonclinical Safety Assessment for Predicting Clinical Adverse Drug Reactions: Correlation Analysis of 142 Approved Drugs in Japan." *The Journal of Toxicological Sciences* 38, no. 4 (2013): 581–98. <https://doi.org/10.2131/jts.38.581>.
53. Burton, Nikolas R., Phillip Kim, and Keriann M. Backus. "Photoaffinity Labelling Strategies for Mapping the Small Molecule–Protein Interactome." *Organic & Biomolecular Chemistry* 19, no. 36 (September 22, 2021): 7792–7809. <https://doi.org/10.1039/D1OB01353J>.
54. Zecha, Jana, Florian P. Bayer, Svenja Wiechmann, Julia Woortman, Nicola Berner, Julian Müller, Annika Schneider, et al. "Decrysting Drug Actions and Protein Modifications by Dose- and Time-Resolved Proteomics." *Science* 380, no. 6640 (April 7, 2023): 93–101. <https://doi.org/10.1126/science.adc3925>.
55. Audagnotto, Martina, and Matteo Dal Peraro. "Protein Post-Translational Modifications: In Silico Prediction Tools and Molecular Modeling." *Computational and Structural Biotechnology Journal* 15 (January 1, 2017): 307–19.

# Supplementary Information

# Embryonic origins of tissues

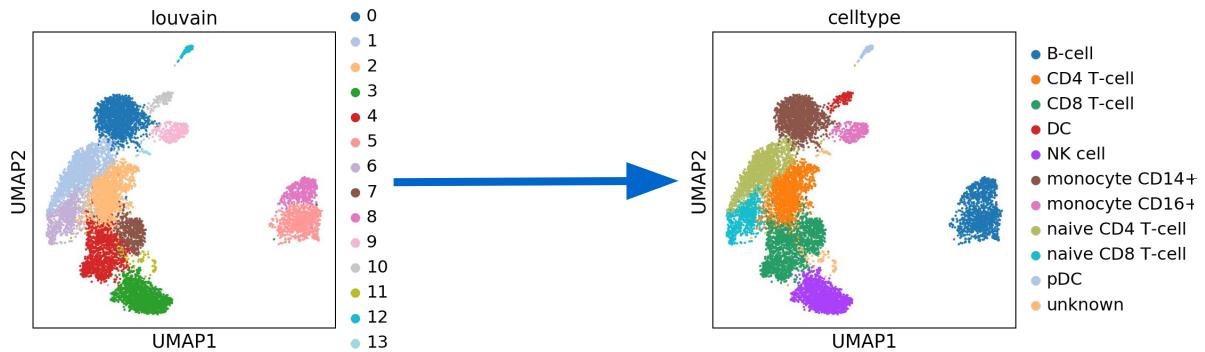
| Germ Layer | Gives rise to:   |  |  |
|------------|--|--|--|
| Ectoderm   | Epidermis, glands on skin, some cranial bones, pituitary and adrenal medulla, the nervous system, the mouth between cheek and gums, the anus                                     |  |  |
| Mesoderm   | Connective tissues proper, bone, cartilage, blood, endothelium of blood vessels, muscle, synovial membranes, serous membranes lining body cavities, kidneys, lining of gonads    |  |  |
| Endoderm   | Lining of airways and digestive system except the mouth and distal part of digestive system (rectum and anal canal); glands (digestive glands, endocrine glands, adrenal cortex) |  |  |

# An intern project: Cell type annotation

From unsupervised clustering and cluster based annotation



Luis Wyss  
RAAN intern 2019



|                 | Gene 1 | Gene 2 | Gene 3 | Gene 4 | Gene 5 | Label      |
|-----------------|--------|--------|--------|--------|--------|------------|
| Training Cell 1 | 10     | 50     | 0      | 12     | 4      | Celltype A |
| Training Cell 2 | 8      | 45     | 78     | 3      | 23     | Celltype B |
| Training Cell 3 | 14     | 55     | 78     | 65     | 55     | Celltype B |
| Training Cell 4 | 78     | 12     | 13     | 9      | 58     | Celltype A |
| Training Cell 5 | 45     | 23     | 65     | 98     | 11     | Celltype C |

To supervised annotation at single-cell level:

|        | Gene 1 | Gene 2 | Gene 3 | Gene 4 | Gene 5 |
|--------|--------|--------|--------|--------|--------|
| Cell 1 | 45     | 45     | 8      | 56     | 3      |
| Cell 2 | 65     | 120    | 78     | 45     | 12     |
| Cell 3 | 79     | 12     | 34     | 65     | 88     |
| Cell 4 | 7      | 59     | 32     | 47     | 62     |

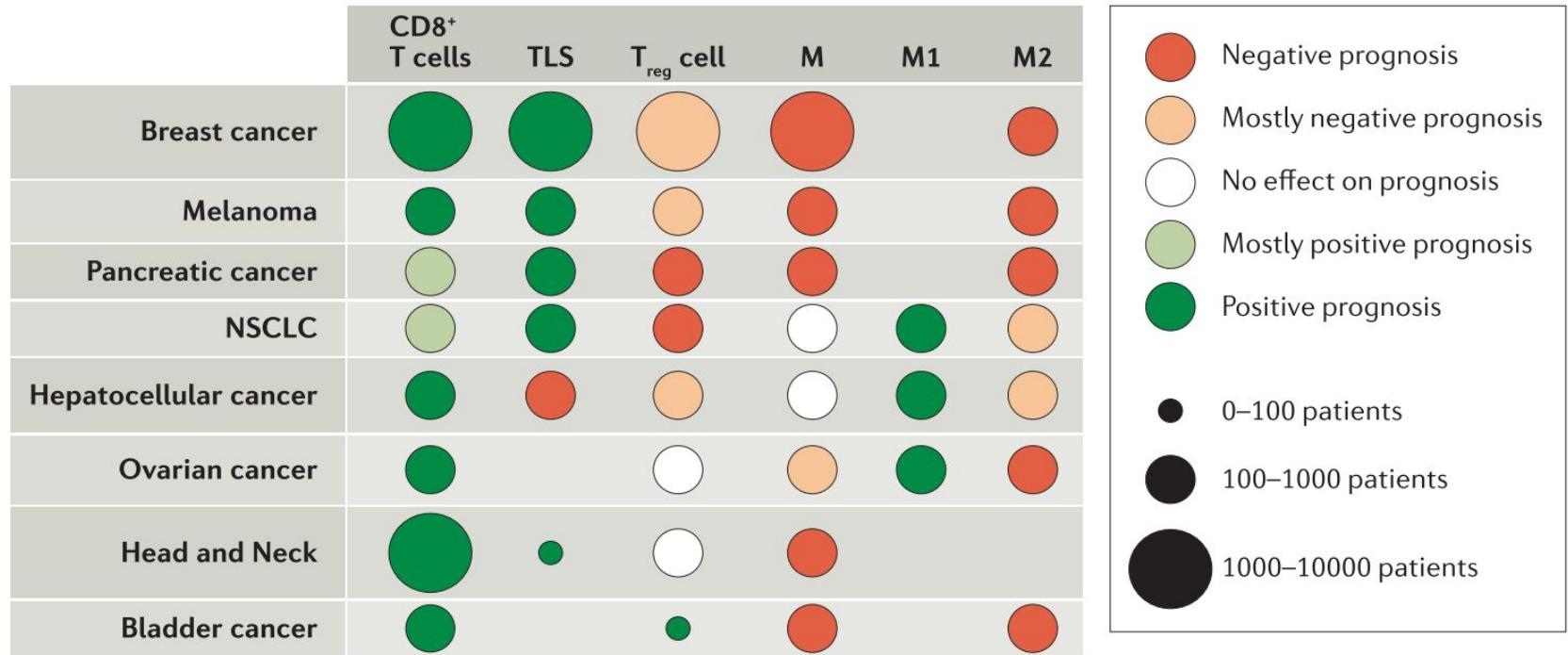


|        | Gene 1 | Gene 2 | Gene 3 | Gene 4 | Gene 5 | Prediction |
|--------|--------|--------|--------|--------|--------|------------|
| Cell 1 | 45     | 45     | 8      | 56     | 3      | Celltype A |
| Cell 2 | 65     | 120    | 78     | 45     | 12     | Celltype B |
| Cell 3 | 79     | 12     | 34     | 65     | 88     | Celltype C |
| Cell 4 | 7      | 59     | 32     | 47     | 62     | Celltype B |



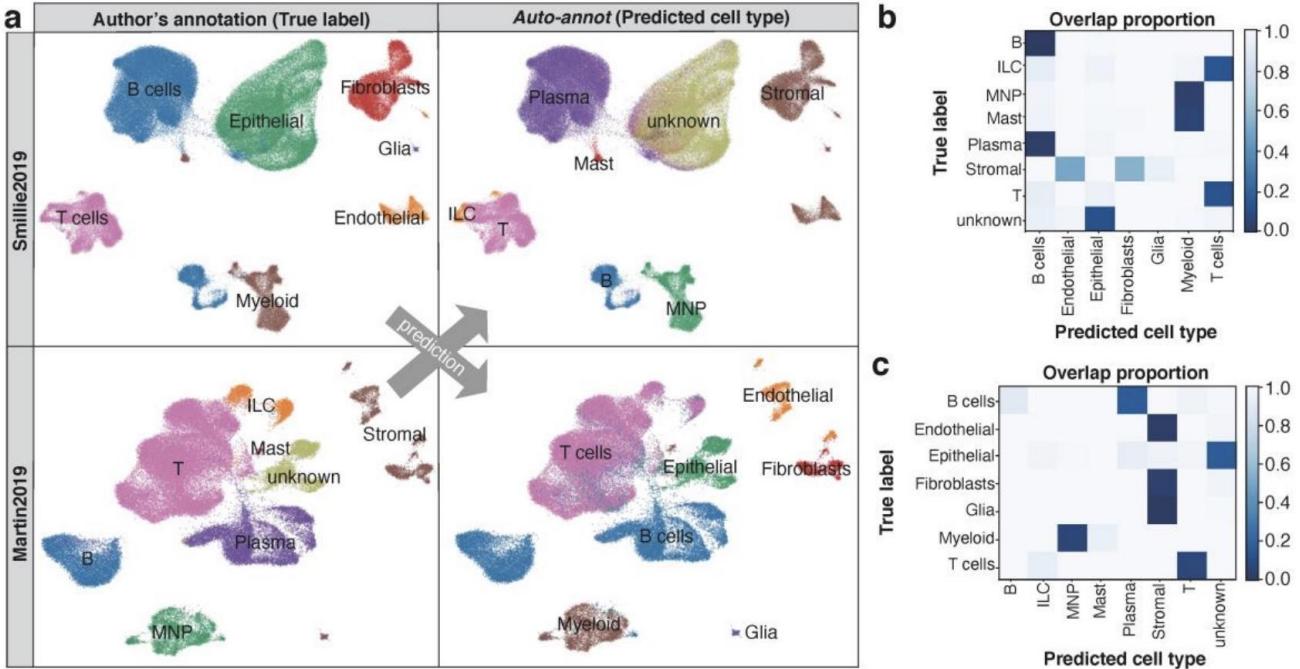
Advantages: (1) automation, (2) annotation independent from clustering, and (3) we can estimate the confidence of prediction

# Abundance of immune cells in tumor microenvironments affect outcome



TLS: tertiary lymphoid structures; T<sub>reg</sub>: regulatory T cells; M: macrophages; M1/M2: subtypes of macrophages

# An example of Inflammatory Bowel Disease (IBD)



We observed Inconsistent cell type nomenclature across studies.  
 Machine learning allows us compare and integrate multiple studies.

# We are living ecosystems

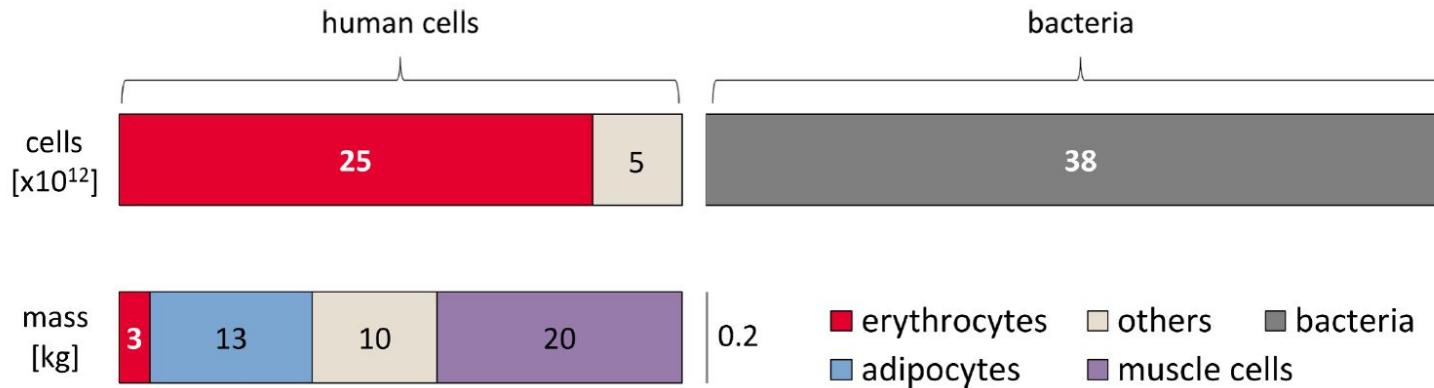
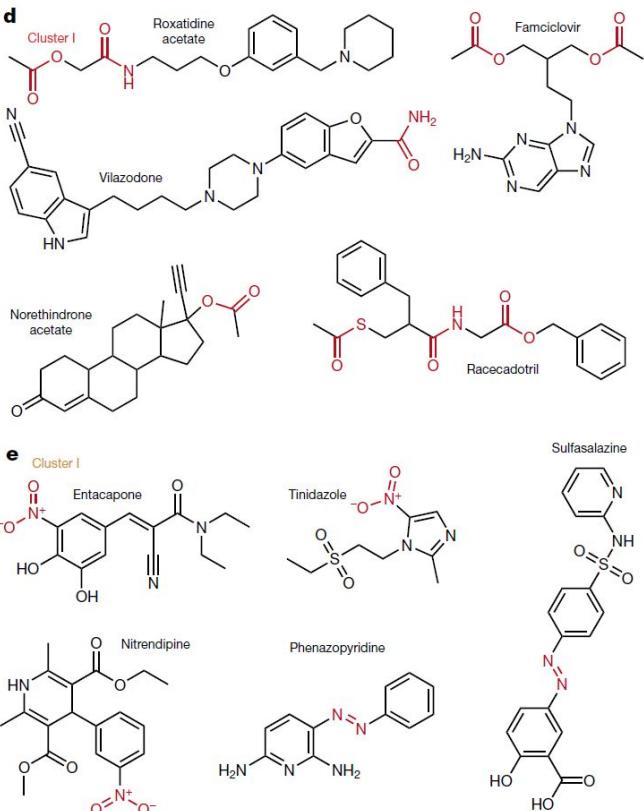
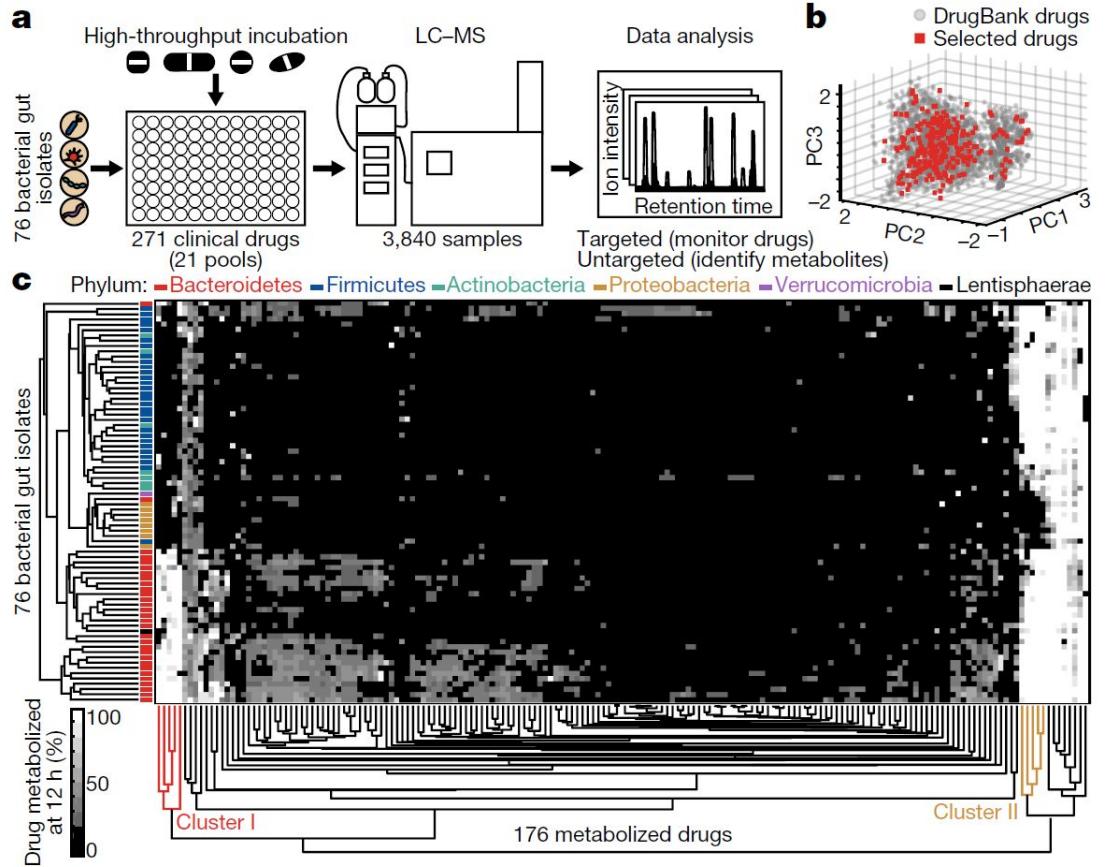


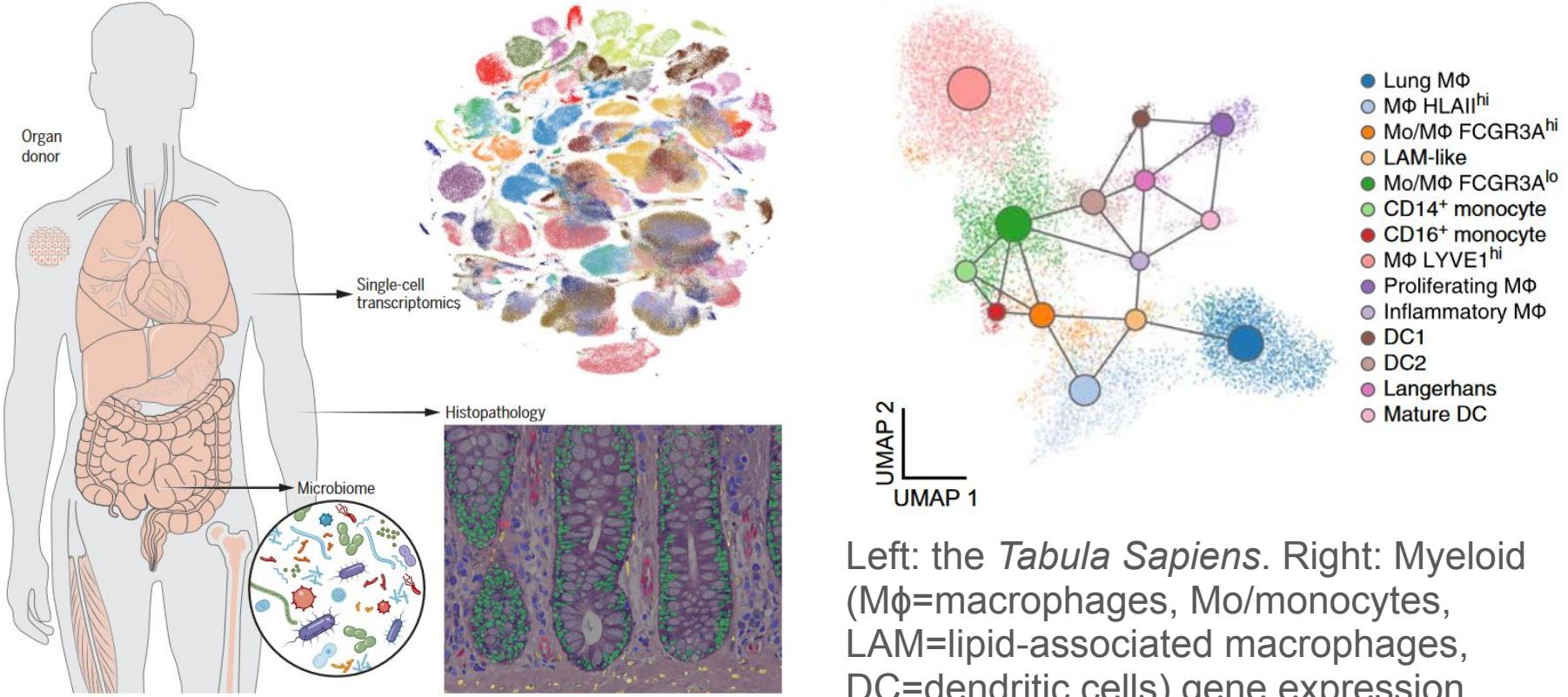
Table 3. B/H ratio for different population. See Table B in [S1 Appendix](#) for full references.

| population segment | body weight [kg] | age [y] | blood volume [L]   | RBC count $[10^{12}/L]$ | colon content [g]  | bac. conc. $[10^{11}/g \text{ wet}]^{(1)}$ | total human cells $[10^{12}]^{(2)}$ | total bacteria $[10^{12}]$ | B:H |
|--------------------|------------------|---------|--------------------|-------------------------|--------------------|--|-------------------------------------|----------------------------|-----|
| ref. man           | 70               | 20–30   | 4.9                | 5.0                     | 420                | 0.92                                       | 30                                  | 38                         | 1.3 |
| ref. woman         | 63               |         | 3.9                | 4.5                     | 480                | 0.92                                       | 21                                  | 44                         | 2.2 |
| young infant       | 4.4              | 4 weeks | 0.4                | 3.8                     | 48                 | 0.92                                       | 1.9                                 | 4.4                        | 2.3 |
| infant             | 9.6              | 1       | 0.8                | 4.5                     | 80                 | 0.92                                       | 4                                   | 7                          | 1.7 |
| elder              | 70               | 66      | 3.8 <sup>(3)</sup> | 4.8                     | 420                | 0.92                                       | 22                                  | 38                         | 1.8 |
| obese              | 140              |         | 6.7                | 5.0 <sup>(4)</sup>      | 610 <sup>(5)</sup> | 0.92                                       | 40                                  | 56                         | 1.4 |

# Gut microbiome can metabolize drugs differently



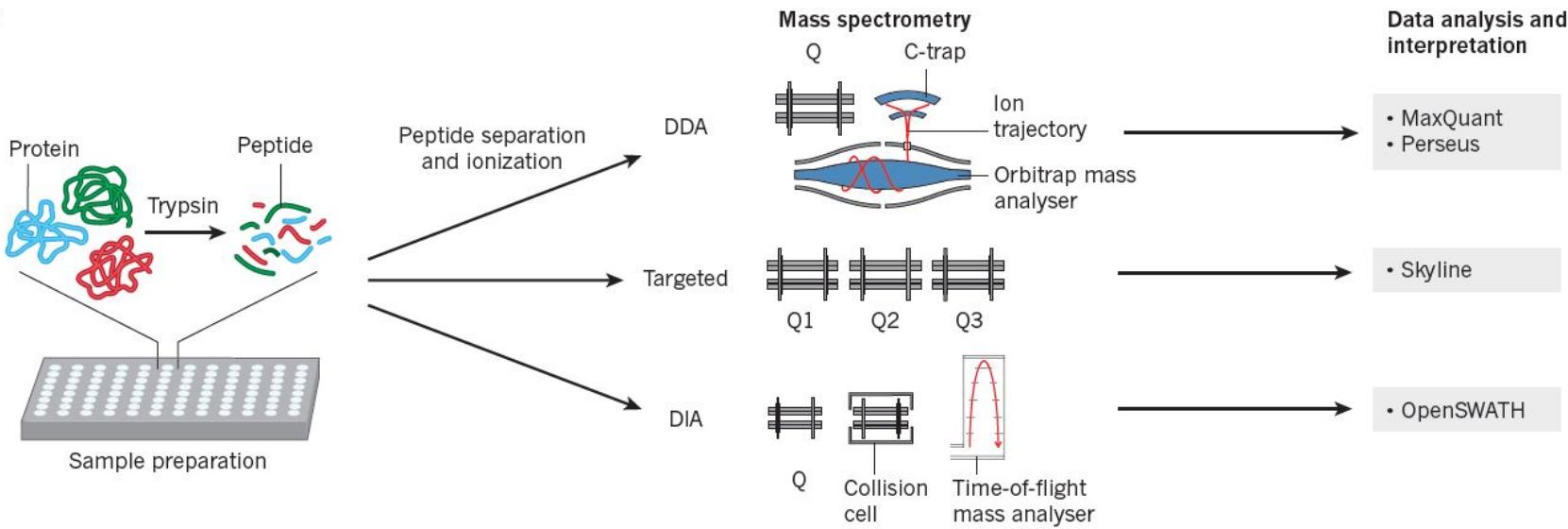
# The *Tabula Sapiens* and other community projects offer reference expression data in healthy donors



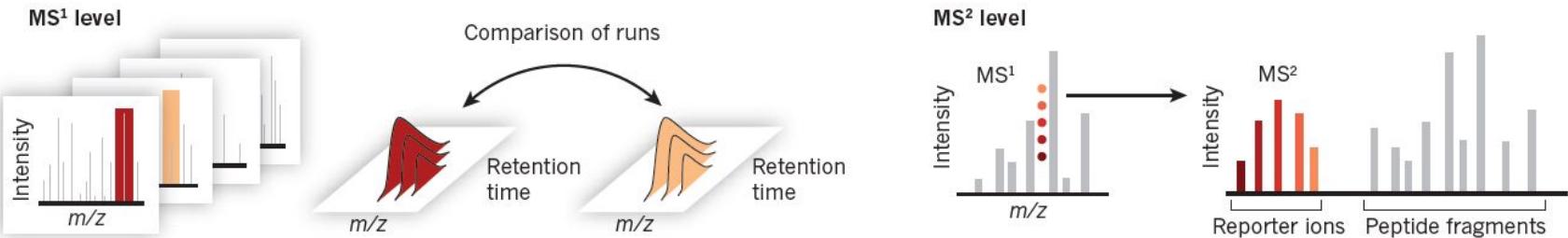
Left: the *Tabula Sapiens*. Right: Myeloid (MΦ=macrophages, Mo/monocytes, LAM=lipid-associated macrophages, DC=dendritic cells) gene expression

# Mass-spectrometry based proteomics

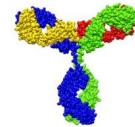
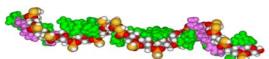
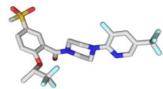
**a**



**b Peptide quantification**



# Comparing modalities with regard to safety assessment



|                          | Small molecules   | Single Stranded Oligos                                    | Biologics   |
|--------------------------|---|---|---|
| Molecular weight         | <1000 D   | 5000-7000 D   | > 30000 D   |
| Manufacture              | Chemical synthesis  | Chemical synthesis  | Biologically-derived                                      |
| Structure                | Single entity, high purity  | Single entity with 10-15% product-related impurities      | Complex, heterogeneous                                    |
| Chemical-driven toxicity | Yes   | Yes   | No  |
| Metabolism               | Species-specific  | Species-independent catabolism by proteolytic degradation | Species-independent catabolism by proteolytic degradation |
| PK                       | Generally short $t_{1/2}$   | Long (tissue) $t_{1/2}$                                   | Long $t_{1/2}$  |
| Some general aspects     | High throughput screening/early safety testing of up to 500 small molecules | Biodistribution with consistent patterns                  | Fewer, yet complex due to biology/immunology              |

# Proteomics enables the elucidation of protein relations in the protein communities

